Problems in Extremal and Probabilistic Combinatorics

by

Lujia Wang
B.S. Nankai University 2011
M.S. Illinois Institute of Technology 2013

Thesis submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Mathematics
in the Graduate College of the
University of Illinois at Chicago, 2018

Chicago, Illinois

Defense Committee:
Dhruv Mubayi, Chair and Advisor
György Turán
Lev Reyzin
Bhaskar DasGupta, Computer Science
Hemanshu Kaul, Illinois Institute of Technology

ACKNOWLEDGMENTS

I want to thank my advisor, Professor Dhruv Mubayi, who guided me through these breath-taking years of learning and developing mathematics. Professor Mubayi is like a church bell, with which a light strike will return you a sublime symphony. His widespread knowledge, his hard work, and his modesty in dealing with the unknown world will keep inspiring me.

I want to thank my parents who always care so much about me.

I thank all my old and new friends who accompanied me along the way.

Many thanks to my thesis committee for taking the time to review my research and give me many useful comments.

CONTRIBUTION OF AUTHORS

Chapter 1 is a literature review that describes the four problems in extremal and probabilistic combinatorics investigated in this thesis together with their backgrounds and significance. Chapter 2 represents a published manuscript (Mubayi, D. and Wang, L. (2018) Multicolour Sunflowers, *Combinatorics, Probability, and Computing,* 1–14. doi:10.1017/S0963548318000160.) My academic advisor, Dr. Dhruv Mubayi and I contributed equally to this research project and the writing of the manuscript. Chapter 3 represents an accepted manuscript (The number of triple systems without even cycles, *Combinatorica*) which is a joint work of mine with Dhruv Mubayi. Chapter 4 represents an unpublished manuscript co-authored by Alan Frieze and Dhruv Mubayi and me. The first two authors motivated the research idea, while Mubayi and I wrote most of the proofs. I also worked out the computational results independently. Chapter 5 represents another unpublished manuscript by Dhruv Mubayi and me. Noticing that another set of authors (71) have developed the same result independently, we decided not to publish it but let them cite us.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

CHAPTER **PAGE**

SUMMARY

We study the following four problems in extremal and probabilistic combinatorics:

1. A sunflower is a collection of distinct sets such that the intersection of any two of them is the same as the common intersection C of all of them, and $|C|$ is smaller than each of the sets. We consider the problems of determining the maximum sum and product of k families of subsets of $[n]$ that contain no sunflower of size k with one set from each family. For the sum, we prove that the maximum is

$$(k-1)2^n + 1 + \sum_{s=0}^{k-2} \binom{n}{s}$$

for all $n \geq k \geq 3$, and for the $k = 3$ case of the product, we prove that the maximum is

$$\left(\frac{1}{8} + o(1)\right) 2^{3n}.$$

We conjecture that for all fixed $k \geq 3$, the maximum product is $(1/8 + o(1))2^{kn}$.

2. For $k \geq 4$, a loose k-cycle C_k is a hypergraph with distinct edges $e_1, e_2, \ldots, e_k$ such that consecutive edges (modulo k) intersect in exactly one vertex and all other pairs of edges are disjoint. Our main result is that for every even integer $k \geq 4$, there exists $c > 0$ such that the number of triple systems with vertex set $[n]$ containing no C_k is at most 2^{cn^2}. An easy construction shows that the exponent is sharp in order of magnitude.

Our proof method is different than that used for most recent results of a similar flavor about enumerating discrete structures, since it does not use hypergraph containers. One novel ingredient is the use of some (new) quantitative estimates for an asymmetric version of the bipartite canonical Ramsey theorem.

3. For $p \in [0, 1]$ and an integer n, let $Q(n, p)$ be the random set system, obtained by picking each subset of $[n]$ independently with probability p. We prove that for many configurations $\mathcal{F}$ that arise naturally in extremal set theory there is a threshold probability p_0 such that if $p \ll p_0$ then asymptotically almost surely $Q(n, p)$ contains no member of $\mathcal{F}$ while if $p \gg p_0$ then asymptotically almost surely $Q(n, p)$ contains many members of $\mathcal{F}$. Our general results imply that $p_0 = (t + 1)^{-n/t}$ is the threshold for the appearance of a matching of size t and is also a threshold for the appearance a chain of size of size t. This generalizes results of Rényi from 1961 who answered a question of Erdős by solving these two problems for $t = 2$. Rényi observed that his approach did not work for larger t for either a matching or chain.

We overcome this problem by using the second moment method on a more restricted class of configurations than the entire family $\mathcal{F}$. Our general result also determines the threshold for the appearance of a sunflower of size t and several other configurations.

4. A family $\mathcal{A} \subset 2^{[n]}$ is intersecting if $A \cap B \neq \emptyset$ for all $A, B \in \mathcal{A}$. We prove that if $p = 2^{-\Theta(\sqrt{n}\log n)}$, the maximum intersecting family in the random set system $Q(n, p)$ is $(1 + o(1))p2^{n-1}$. This is a continuation of the work by Balogh, Bohman and Mubayi who proved the random version of the Erdős–Ko–Rado theorem in 2009. The proof takes advantage

of the hypergraph container method developed independently by Saxton and Thomason, and by Balogh–Morris–Samotij.

CHAPTER 1

INTRODUCTION

1.1 Hypergraphs and families

We first give the definition of hypergraphs together with some relative concepts. Since hypergraphs are the most frequently studied discrete structures in combinatorics, numerous equivalent definitions or synonyms are given under different context. For example, a family of subsets, a design (as in the design theory) and a code (as in the coding theory) may all refer to or equivalent to a hypergraph. In the following chapters, we will switch our terminology for this kind of structures between hypergraphs and families depending on which one is more appropriate to describe certain properties.

Throughout this thesis, we let $[n]$ denote the set $\{1, 2, \dots, n\}$. Let X be a finite set, we denote the power of X by $2^X = \{S : S \subset X\}$, $\binom{X}{r} = \{S \subset X : |S| = r\}$, $\binom{X}{\leq r} = \{S \subset X : |S| \leq r\}$ and $\binom{X}{<r} = \{S \subset X : |S| < r\}$.

A *hypergraph* on the *vertex set* X, H is a collection of subsets of X, i.e. $H \subset 2^X$. The sets contained in a hypergraph are called *edges*. The vertex set X is denoted by $V(H)$. Whenever not specified, it is a convention that the vertex set of a hypergraph H is defined as the union of all its edges:

$$V(H) = \bigcup_{e \in H} e.$$

A subset K of edges of H is called a *sub-hypergraph* of H. Given a subset S of $V(H)$, the *induced sub-hypergraph*, denoted by $H[S]$, is the sub-hypergraph $K \subset H$ where $K = \{e \in H : e \subset S\}$. A hypergraph H is said to be *r-uniform* if $H \subset \binom{X}{r}$, in such a case H is also referred to as an *r-graph*. A *graph* is by convention a 2-graph. The *size* of a hypergraph H is denoted by $|H|$. Given $S \subset V(H)$, the *neighborhood* $N_H(S)$ of S is the set of all $T \subset V(H) \setminus S$ such that $S \cup T \in H$. The *codegree* of S is $d_H(S) = |N_H(S)|$. When the underlying hypergraph is clear from context, we may omit the subscripts in these definitions and write $N(S)$ and $d(S)$ for simplicity. The *sub-edges* of H are the $(r-1)$-subsets of $[n]$ with positive codegree in H. The set of all sub-edges of H is called the *shadow* of H, and is denoted ∂H.

We remark that a hypergraph $H \subset 2^X$ will also be referred to as a *family of subsets of X*, especially in Chapter 2 and Chapter 5. We will introduce equivalent definitions when needed.

1.2 Multicolor Turán problems

1.2.1 Turán problems

Combinatorics focuses on the study of enumeration and properties of discrete structures. One of the most interesting phenomenon in this field of mathematics is that the restrictions on certain local structures induce global changes of the properties of the whole structure. The first example of this phenomenon can be traced back to the very beginning of the graph theory when Leonhard Euler (1) solved the *Seven Bridges Problem of Königsberg* in 1736. Euler proved that to make it possible to traverse all edges of a connected (not necessarily simple) graph–now known as being *eulerian*–it is sufficient to let every vertex have an even degree. The necessity was later proved by Carl Hierholzer (2) in 1871.

Among all the studies of extremal behavior of discrete structures under restrictions, a.k.a *extremal combinatorics*, the Turán type problems which are oriented to graphs and hypergraphs (families) are of central importance. The cornerstone-like result in the area was obtained by Pál Turán (3) in 1941, who proved that an n-vertex graph that does not contain a copy of complete graph on $r+1$ vertices has a maximum of $(1-1/r)n^2/2$ edges which is achieved by complete r-partite graphs with parts of nearly equal size. This result was later extended and generalized in many ways. For example, one may ask the same question of finding the maximum size with other types of forbidden subgraphs or with hypergraph settings.

Let $\mathcal{F}$ be a collection of hypergraphs on $[n]$. Write $\mathrm{Forb}_r(n, \mathcal{F})$ for the set of all r-graphs with vertex set $[n]$ that do not contain a sub-hypergraph isomorphic to a member of $\mathcal{F}$ (henceforth *$\mathcal{F}$-free*). The *Turán number* $\mathrm{ex}_r(n, \mathcal{F})$ is thus defined as

$$\mathrm{ex}_r(n, \mathcal{F}) = \max_{H \in \mathrm{Forb}_r(n, \mathcal{F})} |H|,$$

that is, the maximum number of edges in an r-graph on $[n]$ that is $\mathcal{F}$-free. Notice that for the case that $\mathcal{F}$ contains only one member, say F, we write $\mathrm{Forb}_r(n, F)$ and $\mathrm{ex}_r(n, F)$ for simplicity.

Then, the Turán type problems are to determine (the order of magnitude of) $\mathrm{ex}_r(n, F)$. Now, We turn our attention to the case that $\mathcal{F}$ is the set of all sunflowers, and we generalize the underlying structure to multiple families of subsets of $[n]$, that is, to study the multicolor Turán problem for sunflowers.

1.2.2 Multicolor sunflowers

A *sunflower* (or *strong* Δ*-system*) with k petals is a collection of k sets $\mathcal{S} = \{S_1, \ldots, S_k\}$ such that $S_i \cap S_j = C$ for all $i \neq j$, and $S_i \setminus C \neq \emptyset$ for all $i \in [k]$. The common intersection C is called the *core* of the sunflower and the sets $S_i \setminus C$ are called the *petals*. In 1960, Erdős and Rado (4) proved a fundamental result regarding the existence of sunflowers in a large family of sets of uniform size, which is now referred to as the *sunflower lemma*. It states that if $\mathcal{A}$ is a family of sets of size s with $|\mathcal{A}| > s!(k-1)^s$, then $\mathcal{A}$ contains a sunflower with k petals. Later in 1978, Erdős and Szemerédi (5) gave the following upper bound when the underlying set has n elements.

Theorem 1.2.1 (Erdős, Szemerédi (5))**.** *There exists a constant c such that if $\mathcal{A} \subseteq 2^{[n]}$ with $|\mathcal{A}| > 2^{n-c\sqrt{n}}$ then $\mathcal{A}$ contains a sunflower with 3 petals.*

In the same paper, they conjectured that for n sufficiently large, the maximum number of sets in a family $\mathcal{A} \subseteq 2^{[n]}$ with no sunflowers with three petals is at most $(2-\epsilon)^n$ for some absolute constant $\epsilon > 0$. This conjecture, often referred to as the *weak sunflower lemma*, is closely related to the algorithmic problem of matrix multiplication (6) and remained open for nearly forty years. Recently, this was settled via the polynomial method by Ellenberg and Gijswijt (7) and Croot, Lev and Pach (8) (see also Naslund and Sawin (9)).

A natural way to generalize problems in extremal set theory is to consider versions for multiple families or so-called multicolor or cross-intersecting problems. Beginning with the famous Erdős–Ko–Rado theorem (10), which states that an intersecting family of k-element subsets of $[n]$ has size at most $\binom{n-1}{k-1}$, provided $n \geq 2k$, several generalizations were proved for multiple families that are cross-intersecting. In particular, Hilton (11) showed in 1977 that if t families $\mathcal{A}_1, \ldots, \mathcal{A}_t \subseteq \binom{[n]}{k}$ are cross intersecting (meaning that $A_i \cap A_j \neq \emptyset$ for all $(A_i, A_j) \in \mathcal{A}_i \times \mathcal{A}_j$) and if $n/k \leq t$, then $\sum_{i=1}^{t} |\mathcal{A}_i| \leq t\binom{n-1}{k-1}$. On the other hand, results of Pyber (12) in 1986, that were later slightly refined by Matsumoto and Tokushige (13) and Bey (14), showed that if two families $\mathcal{A} \subseteq \binom{[n]}{k}$, $\mathcal{B} \subseteq \binom{[n]}{l}$ are cross-intersecting and $n \geq \max\{2k, 2l\}$, then $|\mathcal{A}||\mathcal{B}| \leq \binom{n-1}{k-1}\binom{n-1}{l-1}$. These are the first results about bounds on sums and products of the size of cross-intersecting families. More general problems were considered recently, for example for cross t-intersecting families (i.e. pair of sets from distinct families have intersection of size at least t) and r-cross intersecting families (any r sets have a nonempty intersection where each set is picked from a distinct family) and labeled crossing intersecting families, see (15; 16; 17). A more systematic study of multicolored extremal problems (with respect to the sum of the sizes of the families) was initiated by Keevash, Saks, Sudakov and Verstraëte (18), and continued in (19; 20). Cross-intersecting versions of Erdős' problem on weak Δ-systems (for the product of the size of two families) were proved by Frankl and Rödl (21) and by the Mubayi and Rödl (22).

In Chapter 2, we consider multicolor versions of sunflower theorems. Quite surprisingly, these basic questions appear not to have been studied in the literature.

1.3 Enumeration

An important theme in combinatorics is the enumeration of discrete structures that have certain properties. Within extremal combinatorics, people are primarily interested in calculating or to estimating the number of (hyper)graphs that do not contain certain subgraphs. One of the first influential results of this type is the Erdős–Kleitman–Rothschild theorem (23), which implies that the number of triangle-free graphs with vertex set $[n]$ is $2^{n^2/4+o(n^2)}$. This has resulted in a great deal of work on problems about counting the number of graphs with other forbidden subgraphs such as odd cycles (24), complete bipartite graphs (25; 26), octahedron graph (27) and other graphs with certain chromatic properties (28; 29; 30; 31; 32); as well as similar question for other discrete structures (33; 34; 35; 36; 37; 38; 39; 40; 41).

In extremal graph theory, these results show that such problems are closely related to the corresponding extremal problems, more precisely, the *Turán* problems. Recall that $\mathrm{ex}_r(n, F)$ is the maximum number of edges among all r-graphs G on n vertices that contain no copy of F as a (not necessarily induced) subgraph. Henceforth we will call G an F-free r-graph. Write $\mathrm{Forb}_r(n, F)$ for the set of F-free r-graphs with vertex set $[n]$. Since each subgraph of an F-free r-graph is also F-free, we trivially obtain $|\mathrm{Forb}_r(n, F)| \geq 2^{\mathrm{ex}_r(n,F)}$ by taking subgraphs of an F-free r-graph on $[n]$ with the maximum number of edges. On the other hand for fixed r and F,

$$|\mathrm{Forb}_r(n, F)| \leq \sum_{i \leq \mathrm{ex}_r(n,F)} \binom{\binom{n}{r}}{i} = 2^{O(\mathrm{ex}_r(n,F) \log n)},$$

so the issue at hand is the factor $\log n$ in the exponent. The work of Erdős–Kleitman–Rothschild (23) and Erdős–Frankl–Rödl (30) for graphs, and Nagle–Rödl–Schacht (42) for hypergraphs (see also (43) for the case $r = 3$) improves the upper bound above to obtain

$$|\mathrm{Forb}_r(n, F)| = 2^{\mathrm{ex}_r(n,F)+o(n^r)}.$$

Although much work has been done to improve the exponent above (see (44; 28; 29; 27; 31; 45; 32) for graphs and (33; 34; 46; 39; 47; 48) for hypergraphs), this is a somewhat satisfactory state of affairs when $\mathrm{ex}_r(n, F) = \Omega(n^r)$ or F is not r-partite.

In the case of r-partite r-graphs, the corresponding questions appear to be more challenging since the tools used to address the case $\mathrm{ex}_r(n, F) = \Omega(n^r)$ like the regularity lemma are not applicable. The major open problem here when $r = 2$ is to prove that

$$|\mathrm{Forb}_r(n, F)| = 2^{O(\mathrm{ex}_r(n,F))}.$$

The two cases that have received the most attention are for $r = 2$ (graphs) and $F = C_{2l}$ or $F = K_{s,t}$. Classical results of Bondy–Simonovits (49) and Kovári–Sós–Turán (50) yield $\mathrm{ex}_2(n, C_{2l}) = O(n^{1+1/l})$ and $\mathrm{ex}_2(n, K_{s,t}) = O(n^{2-1/s})$ for $2 \le s \le t$, respectively. Although it is widely believed that these upper bounds give the correct order of magnitude, this is not known in all cases. Hence the enumerative results sought in these two cases were

$$|\mathrm{Forb}_2(n, C_{2l})| = 2^{O(n^{1+1/l})} \qquad \text{and} \qquad |\mathrm{Forb}_2(n, K_{s,t})| = 2^{O(n^{2-1/s})}.$$

In 1982, Kleitman and Winston (51) proved that $|\mathrm{Forb}_2(n, C_4)| = 2^{O(n^{3/2})}$ which initiated a 30-year effort on searching for generalizations of the result to complete bipartite graphs and even cycles. Kleitman and Wilson (52) proved similar results for C_6 and C_8 in 1996 by reducing to the C_4 case. Finally, Morris and Saxton (53) recently proved that $|\mathrm{Forb}_2(n, C_{2l})| = 2^{O(n^{1+1/l})}$ and Balogh and Samotij (25; 26) proved that $|\mathrm{Forb}_2(n, K_{s,t})| = 2^{O(n^{2-1/s})}$ for $2 \le s \le t$. Both these results used the hypergraph container method (developed independently by Saxton and Thomason (48), and by Balogh–Morris–Samotij (47)) which has proved to be a very powerful technique in extremal combinatorics. For example, (47) and (48) reproved $|\mathrm{Forb}_r(n, F)| = 2^{\mathrm{ex}_r(n,F)+o(n^r)}$ using containers.

There are very few results in this area when $r > 2$ and $\mathrm{ex}_r(n, F) = o(n^r)$. The only cases solved so far are when F consists of just two edges that intersect in at least t vertices (54), or when F consists of three edges such that the union of the first two is equal to the third (55) (see also (56; 57; 58; 59) for some related results). These are natural points to begin these investigations since the corresponding extremal problems have been studied deeply.

Recently, Kostochka, Mubayi and Verstraëte (60; 61; 62), and independently, Füredi and Jiang (63) (see also (64)) determined the Turán number for several other families of r-graphs including paths, cycles, trees, and expansions of graphs. These hypergraph extremal problems have proved to be quite difficult, and include some longstanding conjectures.

Guided and motivated by these recent developments on the extremal number of hypergraphs, we consider the corresponding enumeration problems focusing on the case of cycles in Chapter 3.

1.4 Threshold functions

The *probabilistic method*, introduced by Paul Erdős in the 1940s, has been proved to be one of the most powerful tools for solving problems in extremal combinatorics. A typical application is to show the existence of certain substructures in a discrete structure of certain size. Usually, a probability space is defined according to the problem setting, and by inequalities derived from the probability theory, one can show there is a positive probability for the existence.

Later the study of the discrete probability spaces itself have gained its own significance. The Erdős–Réyni model of random graph $G(n, p)$–now has become the standard model in most probabilistic combinatorial problems–is the probability space that consists of random graphs generated by taking each possible edge from $\binom{[n]}{2}$ independently with probability p. In their groundbreaking 1960 paper, Erdős and Réyni (65) investigated various asymptotic behaviors of $G(n, p)$–known as to determine the threshold functions–such as the connectivity and the existence of certain subgraphs.

For $p \in [0, 1]$ and an integer n, let $Q(n, p)$ be the random set system, obtained by picking each subset of $[n]$ independently with probability p. We can define threshold function for $Q(n, p)$ in the same fashion of Erdős and Réyni's:

Let P be a property of a realization of the set system $Q(n, p)$, i.e. P is an event that is defined on the probability space $Q(n, p)$. For instance, if $\mathcal{F}$ denotes a family of hypergraphs on $[n]$, one may let P="there is a member of $\mathcal{F}$ that appear as a sub-hypergraph of $Q(n, p)$". $p_0 = p_0(n)$ is called a *threshold function* for a property P if

- when $p = o(p_0), \lim_{n\to\infty} \mathbb{P}(Q(n, p) \models P) = 0$,
- when $p = \omega(p_0), \lim_{n\to\infty} \mathbb{P}(Q(n, p) \models P) = 1$,

or vice versa, where "$Q(n, p) \models P$" means that "$Q(n, p)$ has the property P", or "the event P happens".

A first problem proposed by Erdős about determining threshold functions in $Q(n, p)$ is for the property "there exist two sets that form a chain". In 1961, Rényi (66) proved that the corresponding threshold function is $p_0 = 3^{-n/2}$.

Inspired by these pioneering works, in Chapter 4, we determine the threshold functions for the appearance of a broader class of configurations in the nonuniform random hypergraph $Q(n,p)$. Our work independently reestablishes some results by B. Kreuter (67). We also give some computational results related to certain theorems.

1.5 A randomized Turán problem

Another fruitful topic that lies in the intersection of extremal and probabilistic combinatorics is to solve Turán type problems in random graph or random set system settings. Let $Q(n,p)$ be the (Erdős–Rényi) random set system formed by selecting each subset of $[n]$ independently with probability p, $\mathcal{F}$ be a family of forbidden families. We are interested in the size of the largest $\mathcal{F}$-free family in $Q(n,p)$. However, since sets appear in $Q(n,p)$ at random, one might get a random value when answering the above question. So, similar to the way we treat the threshold functions, we would like to think of the asymptotic behavior of the largest $\mathcal{F}$-free family in $Q(n,p)$. Let $p = p(n)$. Then, $\mathrm{ex}(Q(n,p),\mathcal{F}) = g(n) \pm h(n)$ if

$$\lim_{n\to\infty} \mathbb{P}(\text{the largest } \mathcal{F}\text{-free family in } Q(n,p) \text{ has size } g(n) \pm h(n)) = 1.$$

The problem becomes interesting immediately because the expected result

$$\mathrm{ex}(Q(n,p),\mathcal{F}) = (1+o(1))p \cdot \mathrm{ex}(n,\mathcal{F})$$

does not always hold. For example, if p_0 is the threshold function for the appearance of a member of $\mathcal{F}$, then when $p \ll p_0$ the whole $Q(n,p)$ is asymptotically expected to be $\mathcal{F}$-free, whose size is $p \cdot 2^n$. While, on the opposite extreme, the case with $p = 1$ is identical with the classical Tuán problem which satisfies the expected formula.

1.5.1 Intersecting families in the random set system

A family $\mathcal{A} \subset 2^{[n]}$ is said to be *intersecting* if $A \cap B \neq \emptyset$ for all $A, B \in \mathcal{A}$. It has been well-known that the largest intersecting family in $2^{[n]}$ has size 2^{n-1}. Further, if we restrict $\mathcal{A} \subset \binom{[n]}{k}$ with $k \leq n/2$, then the famous Erdős–Ko–Rado Theorem (10) states that $|\mathcal{A}| \leq \binom{n-1}{k-1}$ if $\mathcal{A}$ is intersecting.

The effort of developing a random version (meaning in the random r-uniform set system $Q_r(n,p)$) of the Erdős–Ko–Rado Theorem with different range of $p(n)$ was initiated by Balogh, Bohman and Mubayi (56) and later improved by Gauy, Hán and Oliverira (68). The work has drawn attention from other authors (54; 59; 58; 69; 70) as well. It seems that the problem is closely related to supersaturation property of the forbidden structures. As long as such a property is established, the invaluable hypergraph container method (independently by Saxton and Thomason (48), and by Balogh–Morris–Samotij (47)) which implies estimations on independent sets in random hypergraphs will take care of the rest.

Let $\mathcal{F} = \{\{A, B\} : A, B \subset [n], A \cap B = \emptyset\}$. In Chapter 5, we determine the value $p = p(n)$ such that $\text{ex}(Q(n,p), \mathcal{F}) = (1 + o(1))p \cdot 2^{n-1}$, and prove that this $p(n)$ is best possible. This can be seen as a continuation of the work by Balogh, Bohman and Mubayi (56) and Gauy, Hán and Oliverira (68). As inspired by the above authors, we solve this problem by proving a supersaturation result regarding intersecting pairs of sets. Then, the main theorem is deduced by the container method.

The very same problem was independently solved by Balogh, Treglown and Wagner (71), where they generalized it to the t-intersecting case.

CHAPTER 2

MULTICOLOR SUNFLOWERS

(Previously published as

Mubayi, D. and Wang, L. (2018) Multicolour Sunflowers, *Combinatorics, Probability, and Computing,* 1–14. doi:10.1017/S0963548318000160.)

The primal problem we want to solve in this chapter is the following: Suppose we have k families (not necessarily uniform) $(\mathcal{A}_i)_{i=1}^k$ of subsets of $[n]$ which together satisfy certain restrictions, what is the "combined size" of these families (hypergraphs)?

Due to the possibility of the appearance of repeated sets in different families, we give the following generalized version of the definition of multicolor sunflowers.

Definition 2.0.1. *Let $A_i \in \mathcal{A}_i \subset 2^{[n]}$ for $i = 1, \ldots, k$. Then, $(A_i)_{i=1}^k$ is a* multicolor (k, s)-sunflower *(or simply a* (k, s)-sunflower*) if*

- *$A_i \cap A_j = C$ for all $1 \leq i < j \leq k$,*
- *there exists an s-set $S \subset [k]$, such that $A_i \setminus C \neq \emptyset$ if and only if $i \in S$. In other words, $A_i = C$ if and only if $i \notin S$.*

$(A_i)_{i=1}^k$ is a multicolor t-sunflower *(or simply a* t-sunflower*) if it is a (k, s)-sunflower for some $s \in [t, k]$, i.e. there are at least t indices i such that $A_i \setminus C \neq \emptyset$. A collection of k families $(\mathcal{A}_i)_{i=1}^k$ is said to be* (k, s)-sunflower-free *(*t-sunflower-free*) if they together do not contain a (k, s)-sunflower (t-sunflower).*

Remark 2.0.2. *We remark that our definition of k-sunflower is identical with the traditional definition of sunflowers used for single family problems. So we will simply refer to it as a* sunflower (with k petals). *If $t < k$, $(A_i)_{i=1}^t$ is a* sunflower with t petals *if each set A_i is in a distinct family $\mathcal{A}_j$ and they form a t-sunflower in these families.*

For any k families that are t-sunflower-free, the problem of upper bounding the size of any single family is uninteresting, since there is no restriction on a particular family. So we are interested in the sum and product of the sizes of these families.

Given integers n and k, let

$$\mathcal{F}(n,k,t) = \{(\mathcal{A}_i)_{i=1}^k : \mathcal{A}_i \subset 2^{[n]} \text{ for } i \in [k] \text{ and } (\mathcal{A}_i)_{i=1}^k \text{ is } t\text{-sunflower-free}\},$$

We define

$$S(n,k,t) := \max_{(\mathcal{A}_i)_{i=1}^k \in \mathcal{F}(n,k,t)} \sum_{i=1}^k |\mathcal{A}_i|,$$

and

$$P(n,k,t) := \max_{(\mathcal{A}_i)_{i=1}^k \in \mathcal{F}(n,k,t)} \prod_{i=1}^k |\mathcal{A}_i|.$$

For the diagonal case, i.e. $t = k$, we simply write $S(n,k)$ and $P(n,k)$.

Our two main results are sharp or nearly sharp estimates on $S(n,k,t)$ and $P(n,3)$. By Theorem 1.2.1 (or (8; 7; 9)) we obtain that

$$S(n,3) \leq 2 \cdot 2^n + 2^{n-c\sqrt{n}}.$$

Indeed, if $|\mathcal{A}|+|\mathcal{B}|+|\mathcal{C}|$ is larger than the RHS above then $|\mathcal{A}\cap\mathcal{B}\cap\mathcal{C}| > 2^{n-c\sqrt{n}}$ by the pigeonhole principle and we find a sunflower in the intersection which contains a sunflower. Our first result removes the last term to obtain an exact result.

Theorem 2.0.3. *For $3 \leq k \leq n$ and $0 \leq t < k$,*

$$S(n,k,t) = (k-1)2^n + \sum_{s=0}^{t-2} \binom{n}{s},$$

and

$$S(n,k) = (k-1)2^n + 1 + \sum_{s=0}^{k-2} \binom{n}{s}.$$

The problem of determining $P(n,k,t)$ seems to be more difficult than that of determining $S(n,k,t)$. Our bounds for general k are quite far apart, but in the case $k = t = 3$ we can refine our argument to obtain an asymptotically tight bound.

Theorem 2.0.4.

$$P(n,3) = \left(\frac{1}{8} + o(1)\right) 2^{3n}.$$

We conjecture that a similar result holds for all $k \geq 3$.

Conjecture 2.0.5. *For each fixed $k \geq 3$,*

$$P(n,k) = \left(\frac{1}{8} + o(1)\right) 2^{kn}.$$

In the next two sections we give the proofs of Theorems 2.0.3 and 2.0.4.

2.1 Sums

In order to prove Theorem 2.0.3, we first deal with s-uniform families and prove a stronger result. Given a sunflower $\mathcal{S} = (A_i)_{i=1}^k$, define its *core size* to be $c(\mathcal{S}) = |C|$, where $C = A_i \cap A_j, i \neq j$. We also need the following notations for the simplicity of presentation.

Definition 2.1.1. *Let there be integers $3 \leq k \leq n$, and families $(\mathcal{A}_i)_{i=1}^k$ of subsets of $[n]$. For each $i \in [k]$ and $s \in [0,n]$, we define*

$$\mathcal{A}_{i,s} = \mathcal{A}_i \cap \binom{[n]}{s} \quad \textit{and} \quad a_s = \sum_{i=1}^k |\mathcal{A}_{i,s}|.$$

Lemma 2.1.2. *Given integers $s \geq 1$, $1 \leq t \leq k$ and $0 \leq c \leq s-1$, let n be an integer such that $n \geq c + t(s-c)$. For $i = 1, \ldots, k$, let $\mathcal{A}_i \subset \binom{[n]}{s}$ such that $(\mathcal{A}_i)_{i=1}^k$ contains no sunflower with t petals and core size c. Then,*

$$a_s \leq \begin{cases} \frac{(t-1)k}{m}\binom{n}{s}, & \textit{if } c + t(s-c) \leq n \leq c + k(s-c) \\ (t-1)\binom{n}{s}, & \textit{if } n \geq c + k(s-c), \end{cases}$$

where $m = \lfloor (n-c)/(s-c) \rfloor$. *Furthermore, these bounds are tight.*

Proof. Randomly take an ordered partition of $[n]$ into $m+2$ parts $X_1, X_2, \ldots, X_{m+2}$ such that $|X_1| = n-(c+m(s-c)), |X_2| = c$, and $|X_i| = s-c$ for $i = 3, \ldots, m+2$, with uniform probability for each partition. For each partition, construct the bipartite graph

$$G = (\{\mathcal{A}_i : i = 1, \ldots, k\} \cup \{X_2 \cup X_j : j \in [3, m+2]\}, E)$$

where a pair $\{\mathcal{A}_i, X_2 \cup X_j\} \in E$ if and only if $X_2 \cup X_j \in \mathcal{A}_i$. If there exists a matching of size t in G, then we will get a sunflower with t petals and core size c (X_2 is the core). This shows that G has matching number at most $t-1$. Then König's theorem implies that the random variable $|E(G)|$ satisfies

$$|E(G)| \leq \begin{cases} (t-1)m, & \text{if } m \geq k \Longleftrightarrow n \geq c + k(s-c), \\ (t-1)k, & \text{if } t \leq m \leq k \Longleftrightarrow c + t(s-c) \leq n \leq c + k(s-c). \end{cases} \tag{2.1}$$

Another way to count the edges of G is through the following expression:

$$|E(G)| = \mathbb{E} \sum_{i=1}^{k} \sum_{j=3}^{k+2} \mathbb{1}_{\{X_2 \cup X_j \in \mathcal{A}_i\}},$$

where $\mathbb{1}_S$ is the characteristic function of the event S. Taking expectations and using the bound Equation 2.1 we obtain

$$|E(G)| = \mathbb{E}\sum_{i=1}^{k}\sum_{j=3}^{k+2}\mathbb{1}_{\{X_2\cup X_j\in\mathcal{A}_i\}}$$
$$\leq \begin{cases} (t-1)m, & \text{if } m\geq k \iff n\geq c+k(s-c), \\ \\ (t-1)k, & \text{if } t\leq m\leq k \iff c+t(s-c)\leq n\leq c+k(s-c). \end{cases}$$

By linearity of expectation,

$$\mathbb{E}\left(\sum_{i=1}^{k}\sum_{j=3}^{m+2}\mathbb{1}_{\{X_2\cup X_j\in\mathcal{A}_i\}}\right) = \sum_{i=1}^{k}\sum_{j=3}^{m+2}\mathbb{P}\left(X_2\cup X_j\in\mathcal{A}_i\right) = \sum_{i=1}^{k}\sum_{j=3}^{m+2}\sum_{A\in\mathcal{A}_i}\mathbb{P}\left(A = X_2\cup X_j\right).$$

Since the partition of $[n]$ is taken uniformly, for any j with $3\leq j\leq m+2$, the set $X_2\cup X_j$ covers all possible s-subsets of $[n]$ with equal probability. Hence for any $A\in\mathcal{A}_i$, we have

$$\mathbb{P}(A = X_2\cup X_j) = \frac{1}{\binom{n}{s}}.$$

So we have

$$\mathbb{E}\left(\sum_{i=1}^{k}\sum_{j=3}^{m+2}\chi_{\{X_2\cup X_j\in\mathcal{A}_i\}}\right) = \sum_{i=1}^{k}\sum_{j=3}^{m+2}\sum_{A\in\mathcal{A}_i}\frac{1}{\binom{n}{s}} = \sum_{i=1}^{k}|\mathcal{A}_i|\frac{m}{\binom{n}{s}} = \frac{m}{\binom{n}{s}}\cdot a_s.$$

Hence by Equation 2.1,

$$a_s \leq \begin{cases} \frac{(t-1)k}{m}\binom{n}{s}, & \text{if } c+t(s-c) \leq n \leq c+k(s-c), \\ \\ (t-1)\binom{n}{s}, & \text{if } n \geq c+k(s-c). \end{cases}$$

Now we are left to show that both upper bounds can be sharp. For the first bound, when $c = 0$, $m = t < k$ and $n = ms$, let each $\mathcal{A}_i$ consist of all s-sets omitting the element 1. A sunflower with $t = m$ petals and core size $c = 0$ is a perfect matching of $[n]$. Since every perfect matching has a set containing 1, there is no sunflower. Clearly $\sum_i |\mathcal{A}_i| = k\binom{n-1}{s} = ((t-1)k/m)\binom{n}{s}$. For the second bound, we can just take $t-1$ copies of $\binom{[n]}{s}$ to achieve equality. $\square$

As a direct application of Lemma 2.1.2, we first prove a bound for sets of size $1 \leq s \leq n-k+1$.

Corollary 2.1.3. *Let integers $3 \leq k \leq n$. For each $0 \leq t \leq k$ and each $1 \leq s \leq n-k+1$. If families $(\mathcal{A}_i)_{i=1}^k$ contain no t-sunflower, then*

$$a_s \leq (k-1)\binom{n}{s}.$$

Proof. Since families $(\mathcal{A}_i)_{i=1}^k$ together contain no t-sunflowers and $t \leq k$, they are also k-sunflower-free. For a given $s \in [1, n-k+1]$, we may take $c = s-1$ to guarantee that

$$c + k(s-c) = s-1+k(s-(s-1)) = k+s-1 \leq n.$$

Then Lemma 2.1.2 implies that

$$a_s \leq (k-1)\binom{n}{s}.$$

$\square$

Lemma 2.1.4. *Let integers $0 \leq t < k$ and $t \leq s \leq k-1$. If $(\mathcal{A}_i)_{i=1}^k$ contains no multicolor t-sunflowers, then*

$$a_{n-s} + (n-s+1)a_{n-s+1} \leq (k-1)(s+1)\binom{n}{n-s}.$$

Proof. Fix any subset B of $[n]$ with $|B| = n-s$, and let $[n] \setminus B = \{x_1, x_2, \ldots, x_s\}$. Consider the k by $s+1$ bipartite graph $G = (V_1 \sqcup V_2, E)$, where $V_1 = \{\mathcal{A}_i\}_{i=1}^k$, $V_2 = \{B\} \cup \{B \cup \{x_i\}\}_{i=1}^s$, and a pair $\{\mathcal{A}_i, A\} \in E$ if $A \in \mathcal{A}_i$. We upper bound the number of edges in G, and claim that $|E(G)| \leq (k-1)(s+1)$.

Suppose the degree of $B \in V_2$, satisfies $d(B) \leq k-s-1$, then we have

$$|E(G)| \leq k-s-1+ks = (k-1)(s+1).$$

Next, suppose the degree $d(B) = k$. Then in the subgraph $H_1 = G[V_1 \sqcup (V_2 \setminus \{B\})]$, there is no matching that saturates $V_2 \setminus \{B\}$. Otherwise we find s families each contains a distinct $(n-s+1)$-set and their pairwise intersection is B; and further all the rest families contain B. This forms an s-sunflower which is also a t-sunflower, contradicting that $(\mathcal{A}_i)_{i=1}^k$ is t-sunflower-free. Hence, we conclude that H_1 has matching number at most $s-1$. Using König's theorem, we get the upper bound

$$|E(G)| \leq k + k(s-1) = ks \leq (k-1)(s+1).$$

Now we may assume $k-s \leq d(B) \leq k-1$. Let $X \subset N(B)$, the neighborhood of B in V_1, such that $|X| = k-s$. Suppose there exists a perfect matching in the subgraph $H_2 = G[(V_1 \setminus X) \sqcup (V_2 \setminus \{B\})]$, then for the similar reason we find an s-sunflower which is also a t-sunflower, contradicting that $(\mathcal{A}_i)_{i=1}^k$ is t-sunflower free. So, H_2 has matching number at most $s-1$. Using König's theorem, we get the upper bound

$$\begin{aligned}|E(G)| &\leq |E(V_1, \{B\})| + |E(X, V_2 \setminus \{B\})| + |E(V_1 \setminus X, V_2 \setminus \{B\})| \\ &\leq k-1 + (k-s)s + (s-1)s = (k-1)(s+1).\end{aligned}$$

Now, consider summing up the number of edges in G (which is built according to the choice of B) over all $B \in \binom{[n]}{k-s}$. In this sum, each $(k-s)$-set is counted once if it appears in some $\mathcal{A}_i$, while each $(k-s+1)$-set, if it appears in some $\mathcal{A}_i$, is counted $k-s+1$ times because there are $\binom{k-s+1}{k-s} = k-s+1$ choices for B that may form such a $(k-s+1)$-set. Therefore we get the inequality,

$$\begin{aligned}
a_{n-s} + (n-s+1)a_{n-s+1} &= \sum_{i=1}^{k} |\mathcal{A}_{i,n-s}| + (n-s+1)\sum_{i=1}^{k} |\mathcal{A}_{i,n-s+1}| \\
&\leq (k-1)(s+1)\binom{n}{n-s}.
\end{aligned}$$

□

We use the lemma above to bound the sum of the number of sets of larger sizes in all families.

Lemma 2.1.5. *Let integers $0 \leq t \leq k$. If $(\mathcal{A}_i)_{i=1}^{k}$ contains no multicolor t-sunflowers, then*

$$\sum_{s=n-k+1}^{n-t+1} a_s \leq (k-1) \sum_{s=n-k+1}^{n-t+1} \binom{n}{s}.$$

Proof. We claim that actually for all $n-k+1 \leq m \leq n-t+1$, we have

$$\sum_{s=n-k+1}^{m} a_s \leq (k-1) \sum_{s=n-k+1}^{m} \binom{n}{s}.$$

We then prove it by induction. The base case $m = n-k+1$ is taken care of by Corollary 2.1.3, i.e. we have

$$a_{n-k+1} \leq (k-1)\binom{n}{n-k+1}. \tag{2.2}$$

Now we may assume $m > n-k+1$ and we have

$$\sum_{s=n-k+1}^{m} a_s \leq (k-1) \sum_{s=n-k+1}^{m} \binom{n}{s}. \tag{2.3}$$

By Lemma 2.1.4,

$$a_m + (m+1)a_{m+1} \leq (k-1)(n-m+1)\binom{n}{m}. \tag{2.4}$$

Take a linear combination of inequalities (Equation 2.2), (Equation 2.3) and (Equation 2.4), and do some simplification on the right hand side, we obtain that

$$\sum_{s=n-k+1}^{m+1} a_s \leq (k-1) \sum_{s=n-k+1}^{m+1} \binom{n}{s}.$$

Thus the proof is complete. □

Now we bound the number of the empty set and singletons for the case $t < k$.

Lemma 2.1.6. *Let integers $0 \le t < k$. If $(\mathcal{A}_i)_{i=1}^k$ contains no multicolor t-sunflowers, then*

$$a_0 + a_1 \le (k-1)(n+1).$$

Proof. Again, by Corollary 2.1.3, we have $a_1 \le (k-1)n$. To prove the combined upper bound, take an arbitrary t-subset $T \subset [n]$ and assume $T = \{x_1, x_2, \ldots, x_t\}$. Consider the k by $t+1$ bipartite graph $G = (V_1 \sqcup V_2, E)$, where $V_1 = \{\mathcal{A}_i\}_{i=1}^k$, $V_2 = \{\emptyset\} \cup \{\{x_i\}\}_{i=1}^t$, and a pair $\{\mathcal{A}_i, A\} \in E$ if $A \in \mathcal{A}_i$. Then by a similar argument as the proof of Lemma 2.1.4, we prove that the number of the graph $|E(G)| \le (k-1)(t+1)$.

When summing up $|E(G)|$ over all possible choices of T, we see that each empty set is counted $\binom{n}{t}$ times while each singleton is counted $\binom{n-1}{t-1}$ times. So we obtain that

$$\binom{n}{t} a_0 + \binom{n-1}{t-1} a_1 \le (k-1)(t+1)\binom{n}{t}.$$

Taking into consideration the inequality $a_1 \le (k-1)n$, we can deduce that

$$a_0 + a_1 \le (k-1)(n+1).$$

□

Now we use this lemma to prove Theorem 2.0.4.

Proof of Theorem 2.0.4. Recall that $n \geq k \geq 3$, $0 \leq t < k$ and we are to show that

$$S(n,k,t) = (k-1)2^n + \sum_{s=0}^{t-2} \binom{n}{s} \tag{2.5}$$

and

$$S(n,k) = S(n,k,k) = (k-1)2^n + 1 + \sum_{s=0}^{k-2} \binom{n}{s}. \tag{2.6}$$

We first show the lower bound by the following examples.

For equation Equation 2.5, let $\mathcal{A}_i = 2^{[n]}$ for $i = 1\ldots,k-1$ and $\mathcal{A}_k = \{S \subset [n] : |S| \geq n-t+2\}$. To see that $(\mathcal{A}_i)_{i=1}^k$ is t-sunflower-free, notice that any t-sunflower uses a set from $\mathcal{A}_k$. So if a set of size at least $n-t+2$ appeared in a t-sunflower, it requires at least $t-1$ other points to form such a sunflower, but then the total number of points in this sunflower is at least $n+1$, a contradiction.

For equation Equation 2.6, let $\mathcal{A}_i = 2^{[n]}$ for $i = 1\ldots,k-1$ and $\mathcal{A}_k = \{\emptyset\} \cup \{S \subset [n] : |S| \geq n-k+2\}$. To see that $(\mathcal{A}_i)_{i=1}^k$ is k-sunflower-free, notice that any k-sunflower uses a set from $\mathcal{A}_k$. The empty set does not lie in any sunflowers. So if a set of size at least $n-k+2$ appeared in a k-sunflower, it requires at least $k-1$ other points to form such a sunflower, but then the total number of points in this sunflower is at least $n+1$, a contradiction.

We then deal with the upper bound for the case that $t = k$. Since $(\mathcal{A}_i)_{i=1}^k$ is k-sunflower-free, for each $s \in [1, n-k+1]$, by Corollary 2.1.3, we have $a_s \leq (k-1)\binom{n}{s}$. Therefore

$$\begin{aligned}
\sum_{i=1}^{k} |\mathcal{A}_i| &= \sum_{i=1}^{k}\sum_{s=0}^{n} |\mathcal{A}_{i,s}| = \sum_{s=0}^{n}\sum_{i=1}^{k} |\mathcal{A}_{i,s}| = \sum_{s=0}^{n} a_s \\
&= a_0 + \sum_{s=1}^{n-k+1} a_s + \sum_{s=n-k+2}^{n} a_s \\
&\le k\binom{n}{0} + \sum_{s=1}^{n-k+1} (k-1)\binom{n}{s} + \sum_{s=n-k+2}^{n} k\binom{n}{s} \\
&\le \sum_{s=0}^{n}(k-1)\binom{n}{s} + \binom{n}{0} + \sum_{s=n-k+2}^{n} \binom{n}{s} \\
&= (k-1)2^n + 1 + \sum_{s=n-k+2}^{n} \binom{n}{s}.
\end{aligned}$$

Now for the case that $0 \le t < k$, first notice that we also have $a_s \le (k-1)\binom{n}{s}$ holds for $s \in [1, n-k+1]$. But further by Lemma 2.1.5 and Lemma 2.1.6, we have

$$\sum_{s=n-k+1}^{n-t+1} a_s \le (k-1) \sum_{s=n-k+1}^{n-t+1} \binom{n}{s}$$

and

$$a_0 + a_1 \le (k-1)(n+1).$$

So a similar calculation is as follows:

$$
\begin{aligned}
\sum_{i=1}^{k} |\mathcal{A}_i| &= \sum_{i=1}^{k} \sum_{s=0}^{n} |\mathcal{A}_{i,s}| = \sum_{s=0}^{n} \sum_{i=1}^{k} |\mathcal{A}_{i,s}| = \sum_{s=0}^{n} a_s \\
&= a_0 + a_1 + \sum_{s=2}^{n-k} a_s + \sum_{s=n-k+1}^{n-t+1} a_s + \sum_{s=n-t+2}^{n} a_s \\
&\leq (k-1)(n+1) + \sum_{s=2}^{n-k} (k-1)\binom{n}{s} + \sum_{s=n-k+1}^{n-t+1} (k-1)\binom{n}{s} + \sum_{s=n-t+2}^{n} k\binom{n}{s} \\
&\leq \sum_{s=0}^{n} (k-1)\binom{n}{s} + \sum_{s=n-t+2}^{n} \binom{n}{s} \\
&= (k-1)2^n + \sum_{s=n-t+2}^{n} \binom{n}{s}.
\end{aligned}
$$

This completes the proof. □

2.2 Products

From the bound on the sum of the families that do not contain a sunflower, we deduce the following bound on the product by using the AM-GM inequality.

Corollary 2.2.1. *Fix $k \geq 3$. As $n \to \infty$,*

$$
\left(\frac{1}{8} + o(1)\right) 2^{kn} \leq P(n,k) \leq \left(\left(\frac{k-1}{k}\right)^k + o(1)\right) 2^{kn}.
$$

Proof. The upper bound follows from Theorem 2.0.4 and the AM-GM inequality,

$$
\prod_{i=1}^{k} |\mathcal{A}_i| \leq \left(\frac{\sum_{i=1}^{k} |\mathcal{A}_i|}{k}\right)^k \leq \left((1+o(1))\frac{(k-1)2^n}{k}\right)^k = (1+o(1))\left(\frac{k-1}{k}\right)^k 2^{kn}.
$$

For the lower bound, we take

$$\mathcal{A}_1 = \mathcal{A}_2 = \{S \subset [n] : 1 \in S\} \cup \{[2, n]\},$$

$$\mathcal{A}_3 = \{S \subset [n] : 1 \notin S\} \cup \{S \subset [n] : |S| \geq n - 1\},$$

and $\mathcal{A}_4 = \mathcal{A}_5 = \ldots = \mathcal{A}_k = 2^{[n]}$. A sunflower with k petals must use three sets from $\mathcal{A}_1, \mathcal{A}_2$, and $\mathcal{A}_3$, call them A_1, A_2, A_3 respectively. These three sets form a sunflower with three petals. If any of these sets is of size at least $n-1$, then it will be impossible to form a 3-petal sunflower with the other two sets. So by their definitions, we have $1 \in A_1 \cap A_2$, but $1 \notin A_3$, which implies $A_1 \cap A_2 \neq A_1 \cap A_3$, a contradiction. So $(\mathcal{A}_i)_{i=1}^k$ is sunflower-free. The sizes of these families are $|\mathcal{A}_1| = |\mathcal{A}_2| = 2^{n-1} + 1, |\mathcal{A}_3| = 2^{n-1} + n$ and $|\mathcal{A}_i| = 2^n$ for $i \geq 4$. Thus,

$$\prod_{i=1}^{k} |\mathcal{A}_i| = \left(\frac{1}{8} + o(1)\right) 2^{kn}$$

as required. □

For any positive integer k we have $(\frac{k-1}{k})^k < 1/e$, so Corollary 2.2.1 implies the upper bound $(1/e + o(1))2^{kn}$ for all $k \geq 3$. For $k = 3$, we will improve the factor in the upper bound from $(2/3)^3 = 0.29629\cdots$ to our conjectured value of 0.125. The main part of our proof is Lemma 2.2.2 below, which proves a much better bound than $S(n, 3) = (2 + o(1))2^n$ for the sum of three sunflower-free families under the assumption that all of them contain a positive proportion of sets.

Lemma 2.2.2. *For all $\epsilon > 0$ there exists $n_0 = n_0(\epsilon) > 0$ such that the following holds for $n > n_0$. Let $\mathcal{A}_i \subset 2^{[n]}$ with $|\mathcal{A}_i| \geq \epsilon 2^n$ for $i \in [3]$, and suppose that $(\mathcal{A}_i)_{i=1}^3$ is sunflower-free. Then*

$$|\mathcal{A}_1| + |\mathcal{A}_2| + |\mathcal{A}_3| \leq \left(\frac{3}{2} + \epsilon\right) 2^n.$$

Lemma 2.2.2 immediately implies Theorem 2.0.4 by the AM-GM inequality as shown below.

Proof of Theorem 2.0.4. Let $\epsilon \in (0, 1/8)$, n_0 be obtained from Lemma 2.2.2 and $n > n_0$. Suppose there is an i, such that $|\mathcal{A}_i| < \epsilon 2^n$. Then

$$\prod_{i=1}^{3} |\mathcal{A}_i| < \epsilon 2^n \cdot 2^n \cdot 2^n < \frac{1}{8} \cdot 2^{3n}.$$

So we may assume that $|\mathcal{A}_i| \geq \epsilon 2^n$ for all i. Thus, by the AM-GM inequality and Lemma 2.2.2,

$$\prod_{i=1}^{3} |\mathcal{A}_i| \leq \left(\frac{|\mathcal{A}_1| + |\mathcal{A}_2| + |\mathcal{A}_3|}{3}\right)^3 \leq \left(\frac{1}{2} + \frac{\epsilon}{3}\right)^3 2^{3n} < \left(\frac{1}{8} + \epsilon\right) 2^{3n}$$

which is the bound sought. □

In the rest of this section we prove Lemma 2.2.2.

2.2.1 Proof of Lemma 2.2.2

We begin with the following lemma, which uses ideas similar to those used in the proof of Lemma 2.1 of (18).

Lemma 2.2.3. *Let $k \geq 3$, $\mathcal{A}_1, \dots \mathcal{A}_k$ be families of subsets of $[n]$ that are sunflower-free. For any real number $\epsilon > 0$, if $|\mathcal{A}_i| \geq \epsilon 2^n$ for all i, then there exists $\delta = \delta(\epsilon) > 0$ and k families $\mathcal{B}_1, \dots, \mathcal{B}_k$ such that the following holds.*

- $|\mathcal{B}_i| \geq \delta 2^n$ *for* $i = 1, \dots, k$,
- $\sum_{i=1}^{k} |\mathcal{A}_i| \leq \sum_{i=1}^{k} |\mathcal{B}_i| + \left(\frac{\epsilon}{2}\right) 2^n$,
- $(\mathcal{B}_i)_{i=1}^{k}$ *is sunflower-free,*
- $\mathcal{B}_1, \dots, \mathcal{B}_k$ *form a laminar system, that is, either* $\mathcal{B}_i \cap \mathcal{B}_j = \emptyset$, $\mathcal{B}_i \subset \mathcal{B}_j$, *or* $\mathcal{B}_j \subset \mathcal{B}_i$ *for all* $i \neq j$.

Proof. The families $\mathcal{A}_i$, $i = 1, \dots, k$ form a collection of subsets of $2^{[n]}$, hence they induce a partition of $2^{[n]}$ into at most 2^k parts. More precisely, the disjoint parts (some may be empty) in this partition are

$$\mathcal{X}_I = \bigcap_{i \in I} \mathcal{A}_i \bigcap_{i \in [k] \setminus I} \mathcal{A}_i^c, \text{ where } I \subset [k].$$

Take $\delta = \epsilon/(k 2^k)$. For each $I \subset [k]$, if $|\mathcal{X}_I| < \delta 2^n$, update the $\mathcal{A}_i$s by deleting $\mathcal{X}_I$ from all $\mathcal{A}_i$s that contain it, that is, all $\mathcal{A}_i$s with $i \in I$. At the end of this process, let the resulting families be $\mathcal{A}'_i, i = 1 \dots, k$. Now, all $\mathcal{X}_I$s that are nonempty have size at least $\delta 2^n$. For each original $\mathcal{A}_i$, the families $\mathcal{A}_i \cap \mathcal{A}_j, j \in [k] \setminus \{i\}$ induce a partition on it into at most 2^{k-1} parts. So, after the above deletion steps the remaining $\mathcal{A}'_i$ has size at least

$$|\mathcal{A}'_i| \geq \epsilon 2^n - 2^{k-1} \delta 2^n = \epsilon 2^n - 2^{k-1} \frac{\epsilon}{k 2^k} 2^n = \left(1 - \frac{1}{2k}\right) \epsilon 2^n.$$

If $\mathcal{X}_I < \delta 2^n$, it is deleted from all $|I|$ of the $\mathcal{A}_i$s that contain it. Hence, the total number of deleted parts with repetition is at most

$$\sum_{i=1}^{n} i\binom{k}{i} = k2^{k-1}.$$

So, we have

$$\sum_{i=1}^{k} |\mathcal{A}_i| \leq \sum_{i=1}^{k} |\mathcal{A}'_i| + k2^{k-1}\delta 2^n = \sum_{i=1}^{k} |\mathcal{A}'_i| + \left(\frac{\epsilon}{2}\right) 2^n.$$

Two families $\mathcal{A}$ and $\mathcal{B}$ are said to be *crossing* if all three of $\mathcal{A}\cap\mathcal{B}$, $\mathcal{A}\setminus\mathcal{B}$ and $\mathcal{B}\setminus\mathcal{A}$ are nonempty. For each pair of crossing families $\mathcal{A}'_i$ and $\mathcal{A}'_j$, replace $\mathcal{A}'_i$ and $\mathcal{A}'_j$ by $\mathcal{A}'_i \cap \mathcal{A}'_j$ and $\mathcal{A}'_i \cup \mathcal{A}'_j$. Call the resulting families $\mathcal{B}_1, \ldots, \mathcal{B}_k$.

Notice first that at the end of the process (which terminates after at most $\binom{k}{2}$ steps, because it increases the number of inclusion related pairs at every step), the families $\mathcal{B}_1, \ldots, \mathcal{B}_k$ contain no crossing pairs, hence form a laminar system. Secondly, the sum of the sizes of the families remains the same, since $|X \cap Y| + |X \cup Y| = |X| + |Y|$ for all sets X, Y. Hence we get

$$\sum_{i=1}^{k} |\mathcal{A}_i| \leq \sum_{i=1}^{k} |\mathcal{A}'_i| + \left(\frac{\epsilon}{2}\right) 2^n = \sum_{i=1}^{k} |\mathcal{B}_i| + \left(\frac{\epsilon}{2}\right) 2^n.$$

Next, notice that all parts of the partition induced by $\mathcal{A}'_i, i = 1, \ldots, k$ have size at least $\delta 2^n$. Moreover, the steps of replacing two crossing families by their intersection and union only create new families that consists of the union of nonempty parts. This yields that $|\mathcal{B}_i| \geq \delta 2^n$ for all $i \in [k]$. Finally, we claim that $(\mathcal{B}_i)_{i=1}^k$ is sunflower-free. The families $(\mathcal{A}'_i)_{i=1}^k$ are certainly sunflower-free because $\mathcal{A}_i' \subset \mathcal{A}_i$ for all i and $(\mathcal{A}_i)_{i=1}^k$ is sunflower-free. So we are left to show that the steps of removing crossing pairs do not introduce sunflowers. Suppose we have families $(\mathcal{C}_i)_{i=1}^k$, w.l.o.g, the crossing pair $\mathcal{C}_1, \mathcal{C}_2$ are replaced by $\mathcal{C}_1 \cap \mathcal{C}_2$ and $\mathcal{C}_1 \cup \mathcal{C}_2$, and suppose that $C_i, i = 1, \ldots, k$ with $C_1 \in \mathcal{C}_1 \cap \mathcal{C}_2$, $C_2 \in \mathcal{C}_1 \cup \mathcal{C}_2$ and $C_i \in \mathcal{C}_i, i \geq 3$ is a sunflower in the resulting families. Then, w.l.o.g, C_2 is in $\mathcal{C}_2$. Thus we find that $C_i, i = 1, \ldots, k$ is also a sunflower in $(\mathcal{C}_i)_{i=1}^k$. This completes the proof. $\square$

We will use the following lemma which follows from well-known properties of binomial coefficients (we omit the standard proofs).

Lemma 2.2.4. *For each $\delta > 0$, there exists a real number $\alpha = \alpha(\delta)$ and integer n_0 such that for $n > n_0$, every family $\mathcal{A}$ of subsets of $[n]$ with size $|\mathcal{A}| \geq \delta 2^n$ contains a set S with $|S| \in [n/2 - \alpha\sqrt{n}, n/2 + \alpha\sqrt{n}]$. Further, for each $\gamma \in (0, \delta)$, there exists a $\beta = \beta(\gamma)$, such that all but at most $\gamma 2^n$ elements in $\mathcal{A}$ have size in $[n/2 - \beta\sqrt{n}, n/2 + \beta\sqrt{n}]$.*

Now we have all the necessary ingredients to prove Lemma 2.2.2.

Proof of Lemma 2.2.2. Let $\delta = \epsilon/(3 \cdot 2^3) = \epsilon/(24)$ as in the proof of Lemma 2.2.3. By Theorem 1.2.1, we have $|\mathcal{A}_1 \cap \mathcal{A}_2 \cap \mathcal{A}_3| \leq 2^{n - c\sqrt{n}} < \delta 2^n$ for large enough n. Apply Lemma 2.2.3 to obtain families $\mathcal{B}_i, i = 1, 2, 3$ such that

- $|\mathcal{B}_i| \geq \delta 2^n$ for $i = 1, 2, 3$,
- $\sum_{i=1}^3 |\mathcal{A}_i| \leq \sum_{i=1}^3 |\mathcal{B}_i| + \left(\frac{\epsilon}{2}\right) 2^n$,
- $(\mathcal{B}_1, \mathcal{B}_2, \mathcal{B}_3)$ is sunflower-free,
- $\mathcal{B}_1, \mathcal{B}_2, \mathcal{B}_3$ form a laminar system.

Moreover, since $|\mathcal{A}_1 \cap \mathcal{A}_2 \cap \mathcal{A}_3| < \delta 2^n$, the intersection of all three families is deleted from all three of them in the process of forming $\mathcal{B}_i$s which yields $\mathcal{B}_1 \cap \mathcal{B}_2 \cap \mathcal{B}_3 = \emptyset$. The rest of the proof is devoted to showing the claim below.

Claim 2.2.5.

$$|\mathcal{B}_1| + |\mathcal{B}_2| + |\mathcal{B}_3| \leq \left(\frac{3}{2} + \frac{\epsilon}{2}\right) 2^n.$$

Proof. The laminar system formed by the three families with an empty common intersection falls into the following three types. Let $\{\mathcal{A}, \mathcal{B}, \mathcal{C}\} = \{\mathcal{B}_1, \mathcal{B}_2, \mathcal{B}_3\}$ and $a := |\mathcal{A}|, b := |\mathcal{B}|$, and $c := |\mathcal{C}|$.

Case I. $\mathcal{A}, B, C$ are mutually disjoint. In this case, trivially we have $a + b + c \leq 2^n$ which is even better than what we need.

Case II. $\mathcal{A} \supset \mathcal{B}$ and $\mathcal{A} \cap \mathcal{C} = \emptyset$. Since $|\mathcal{C}| \geq \delta 2^n$, we may pick an $S \in \mathcal{C}$ with $|S| \in [n/2 - \alpha\sqrt{n}, n/2 + \alpha\sqrt{n}]$ by Lemma 2.2.4. Now for each subset $T \subset S$, consider the subfamily of $\mathcal{B}$ defined by

$$\mathcal{B}_T = \{B \in \mathcal{B} : B \cap S = T\}.$$

Clearly, these subfamilies form a partition of $\mathcal{B}$, i.e. $\mathcal{B} = \bigsqcup_{T \subset S} \mathcal{B}_T$. Now we define a new family derived from $\mathcal{B}'_T$

$$\mathcal{B}'_T = \{B \setminus T : B \in \mathcal{B}_T\}.$$

There is a naturally defined bijection between $\mathcal{B}_T$ and $\mathcal{B}'_T$, so $|\mathcal{B}_T| = |\mathcal{B}'_T|$. **Claim.** $b \leq (1 + \frac{\epsilon}{2})2^{n-1}$.

Proof. We first show that $\mathcal{B}'_T \setminus \{\emptyset\}$ is an intersecting family if $T \subsetneq S$. Indeed, suppose there are disjoint nonempty sets $B_1, B_2 \in \mathcal{B}'_T$, then we find a sunflower consisting of $B_1 \cup T \in \mathcal{B} \subset \mathcal{A}$, $B_2 \cup T \in \mathcal{B}$ and $S \in \mathcal{C}$. So $|\mathcal{B}'_T| \le 2^{n-|S|-1} + 1$, which yields the following upper bound for $|\mathcal{B}|$:

$$
\begin{aligned}
b &= \sum_{T \subset S} |\mathcal{B}_T| = \sum_{T \subset S} |\mathcal{B}'_T| = \sum_{T \subsetneq S} |\mathcal{B}'_T| + |\mathcal{B}'_S| \\
&\le (2^{|S|} - 1)(2^{n-|S|-1} + 1) + 2^{n-|S|} = 2^{n-1} + 2^{|S|} - 2^{n-|S|-1} - 1 + 2^{n-|S|} \\
&\le 2^{n-1} + 2^{n/2+\alpha\sqrt{n}} - 2^{n-(n/2+\alpha\sqrt{n})-1} + 2^{n-(n/2-\alpha\sqrt{n})} \\
&= 2^{n-1} + 2^{n/2+\alpha\sqrt{n}} - 2^{n/2-\alpha\sqrt{n}-1} + 2^{n/2+\alpha\sqrt{n}} \\
&\le \left(1 + \frac{\epsilon}{2}\right) 2^{n-1},
\end{aligned}
$$

where the last inequality holds for large enough n. $\square$

Since $\mathcal{A} \cap \mathcal{C} = \emptyset$, the Claim implies that

$$
a + b + c \le 2^n + \left(1 + \frac{\epsilon}{2}\right) 2^{n-1} = \left(\frac{3}{2} + \frac{\epsilon}{2}\right) 2^n.
$$

Case III. $\mathcal{A} \supset (\mathcal{B} \cup \mathcal{C})$ and $\mathcal{B} \cap \mathcal{C} = \emptyset$. We first fix $\gamma = \min\{\delta, \epsilon/12\}$, find $\beta = \beta(\gamma)$ as in Lemma 2.2.4. Then all but at most $\gamma 2^n \le (\epsilon/12)2^n$ sets in each family are of size in $[n/2 - \beta\sqrt{n}, n/2 + \beta\sqrt{n}]$. Hence we have

$$
a + b + c \le |\mathcal{A}_\beta| + |\mathcal{B}_\beta| + |\mathcal{C}_\beta| + \frac{\epsilon}{4} \cdot 2^n,
$$

where $\mathcal{F}_\beta = \{F \in \mathcal{F} : n/2 - \beta\sqrt{n} \leq |F| \leq n/2 + \beta\sqrt{n}\}$. It remains to show that

$$|\mathcal{A}_\beta| + |\mathcal{B}_\beta| + |\mathcal{C}_\beta| \leq \left(\frac{3}{2} + \frac{\epsilon}{4}\right) 2^n.$$

We may assume $\mathcal{A} = \mathcal{A}_\beta, \mathcal{B} = \mathcal{B}_\beta$ and $\mathcal{C} = \mathcal{C}_\beta$; our task is to prove $a + b + c \leq (3/2 + \epsilon/4)\, 2^n$. Consider a pair of sets $(B, C) \in \mathcal{B} \times \mathcal{C}$ which satisfies the following two conditions:

- $B \cup C \neq [n]$,
- $B \setminus C \neq \emptyset$ and $C \setminus B \neq \emptyset$.

Let $A = \overline{B \triangle C} = (B \cap C) \cup \overline{B \cup C}$. Then $A \notin \mathcal{A}$, otherwise A, B, C together form a sunflower. Hence the number of such As is at most $2^n - a$. We claim that for each such A, there are at most $(1 + \epsilon/4)2^{n-1}$ pairs $(B, C) \in \mathcal{B} \times \mathcal{C}$ with the two properties above such that $A = \overline{B \triangle C}$. Indeed, for a given A, we first partition it into two ordered parts X_1, X_2 with $X_2 \neq \emptyset$ (here X_2 corresponds to $\overline{B \cup C}$). There are $2^{|A|} - 1$ ways to do so. Next we count the number of such pairs (B, C) such that $B \cap C = X_1$ and $\overline{B \cup C} = X_2$. This number at most $1/2$ of the number of ordered partitions of $[n] \setminus A$ into two nonempty parts. The ratio $1/2$ comes from the fact that for each ordered bipartition $[n] \setminus A = X_3 \sqcup X_4$, if $(X_3 \cup X_1, X_4 \cup X_1) \in (\mathcal{B} \times \mathcal{C})$, then we cannot also have $(X_4 \cup X_1, X_3 \cup X_1) \in (\mathcal{B} \times \mathcal{C})$, because $\mathcal{B}$ and $\mathcal{C}$ are disjoint. So only half of the ordered bipartitions could actually become desired pairs. Consequently, the number of such pairs (B, C) is $(2^{n-|A|} - 2)/2 = 2^{n-|A|-1} - 1$. The total number (B, C) that give the same A is therefore at most

$$
\begin{aligned}
(2^{|A|} - 1)(2^{n-|A|-1} - 1) &= 2^{n-1} - 2^{|A|} - 2^{n-|A|-1} + 1 \\
&\leq 2^{n-1} - 2^{n/2-\beta\sqrt{n}} - 2^{n-(n/2+\beta\sqrt{n})-1} + 1 \\
&\leq \left(1 + \frac{\epsilon}{4}\right) 2^{n-1}.
\end{aligned}
$$

Here we use the assumption that $\mathcal{A} = \mathcal{A}_\beta$, which implies $|A| \in [n/2 - \beta\sqrt{n}, n/2 + \beta\sqrt{n}]$, and n is large enough. This yields

$$bc \leq (2^n - a)\left(1 + \frac{\epsilon}{4}\right) 2^{n-1} + 3^{n+1}$$

where the error term 3^{n+1} arises from the number of pairs $(B, C) \in \mathcal{B} \times \mathcal{C}$ such that either $B \cup C = [n]$, $B \subset C$ or $C \subset B$. If $(2^n - a)(\epsilon/4)\, 2^{n-1} < 3^{n+1}$, then $bc < (10/\epsilon)3^{n+1}$ and this contradicts $b, c \geq \delta 2^n$. Therefore

$$bc \leq (2^n - a)\left(1 + \frac{\epsilon}{4}\right) 2^{n-1} + 3^{n+1} \leq (2^n - a)\left(1 + \frac{\epsilon}{2}\right) 2^{n-1}.$$

Consequently, we have

$$a \leq 2^n - \frac{bc}{(1 + \epsilon/2)2^{n-1}}.$$

By the same argument used for the proof of the Claim in Case II, we can show that $b \le (1+\epsilon/2)2^{n-1}$ and $c \le (1+\epsilon/2)2^{n-1}$. Now we obtain

$$a+b+c \le 2^n - \frac{bc}{(1+\epsilon/2)2^{n-1}} + b + c = f(b,c) \le \left(\frac{3}{2}+\frac{\epsilon}{4}\right)2^n$$

where the last inequality follows by maximizing the function $f(b,c)$ subject to the constraints $b, c \in I = [\delta 2^n, (1+\epsilon/2)2^{n-1}]$. Indeed, setting $\partial_b f = \partial_c f = 0$ we conclude that the extreme points occur at the boundary of $I \times I$. In fact, the maximum is achieved at $(b,c) = ((1+\epsilon/2)2^n, (1+\epsilon/2)2^n)$, and $f((1+\epsilon/2)2^n, (1+\epsilon/2)2^n) = (3/2+\epsilon/4)2^n$ as claimed above. $\square$

2.3 Concluding remarks

- By the monotonicity of the function $P(n,k,t)$ in t, Theorem 2.0.4 implies for each fixed $0 \le t \le 3$,

$$P(n,3,t) = \left(\frac{1}{8}+o(1)\right)2^{3n}.$$

The case $t=0$ is particularly interesting. Let $P^*(n,k) = P(n,k,0)$, $p^*(n,k) = P^*(n,k)/2^{kn}$ and $p(n,k) = P(n,k)/2^{kn}$. As pointed out by a referee, it is easy to show that $p^*(n,k)$ is monotone increasing as a function of n for each fixed $k \ge 3$, while $p(n,k)$ is not. Indeed, given a collection of optimal families $(\mathcal{A}_i)_{i=1}^k$ for $P^*(n,k)$, we can construct k families of subsets of $[n+1]$ that are 0-sunflower-free with the product of their sizes at least $2^k P^*(n,k)$ as follows. We "double" each $\mathcal{A}_i$ in the following way to get new families:

$$\mathcal{B}_i = \mathcal{A}_i \cup \{A \cup \{n+1\} : A \in \mathcal{A}_i\}, \quad i \in [k].$$

Clearly, $\prod_{i=1}^{k} |\mathcal{B}_i| = \prod_{i=1}^{k} 2|\mathcal{A}_i| = 2^k P^*(n,k)$ and it is an easy exercise to show that $(\mathcal{B}_i)_{i=1}^{k}$ contains no 0-sunflower. Since $p^*(n,k) \leq 1$, we conclude that $p^*(k) := \lim_{n\to\infty} p^*(n,k)$ exists. Clearly $p^*(3) = 1/8$, and in general $1/8 \leq p^*(k) \leq (1-1/k)^k < 1/e$. Further, for a fixed $k \geq 4$, if one can show that there exists a single value n_0 such that $p^*(n_0,k) > 1/8$, then by the monotonicity of $p^*(n,k)$ and $P^*(n,k) \leq P(n,k)$, Conjecture 3.0.3 would be disproved.

- Our approach for $S(n,k)$ is simply to average over a suitable family of partitions. It can be applied to a variety of other extremal problems, for example, it yields some results about cross intersecting families proved by Borg (72). It also applies to the situation when the number of colors is more than the size of the forbidden configuration. In particular, the proof of Lemma 2.0.3 yields the following more general statement.

- Another general approach that applies to the sum of the sizes of families was initiated by Keevash–Saks–Sudakov–Verstraëte (18). We used the idea behind this approach in Lemma 2.2.2. Both methods can be used to solve certain problems. For example, as pointed out to us by Benny Sudakov, the approach in (18) can be used to prove the $k = 3$ case of Theorem 2.0.4 (and perhaps other cases too).

CHAPTER 3

THE NUMBER OF TRIPLE SYSTEMS WITHOUT EVEN CYCLES

In this chapter, we prove results on the number of hypergraph without loose cycles.

Definition 3.0.1. *For each integer $k \geq 3$, a k-cycle C_k is a hypergraph with distinct edges $e_1, \ldots, e_k$ and distinct vertices $v_1, \ldots, v_k$ such that $e_i \cap e_{i+1} = \{v_i\}$ for all $1 \leq i \leq k-1$, $e_1 \cap e_k = \{v_k\}$ and $e_i \cap e_j = \emptyset$ for all other pairs i, j.*

Sometimes we refer to C_k as a *loose* or *linear* cycle. To simplify notation, we will omit the parameter r when the cycle C_k is a subgraph of an r-graph.

Since $\mathrm{ex}_r(n, C_k) = O(n^{r-1})$, we obtain the upper bound

$$|\mathrm{Forb}_r(n, C_k)| = 2^{O(n^{r-1} \log n)}$$

when r and k are fixed and $n \to \infty$. Our main result is the following theorem, which improves this upper bound.

Theorem 3.0.2. *For every even integer $k \geq 4$, there exists $c = c(k)$, such that*

$$|\mathrm{Forb}_3(n, C_k)| < 2^{c\,n^2}.$$

Since trivially $\mathrm{ex}_r(n, C_k) = \Omega(n^{r-1})$ for all $r \geq 3$, we obtain $|\mathrm{Forb}_3(n, C_k)| = 2^{\Theta(n^2)}$ when k is even. We conjecture that a similar result holds for $r > 3$ and cycles of odd length.

Conjecture 3.0.3. [1] *For fixed $r \geq 3$ and $k \geq 3$ we have* $|\mathrm{Forb}_r(n, C_k)| = 2^{\Theta(n^{r-1})}$.

Almost all recent developments in this area have relied on the method of hypergraph containers that we mentioned above. What is perhaps surprising about the current work is that the proofs do not use hypergraph containers. Instead, our methods employ old and new tools in extremal (hyper)graph theory. The old tools include the extremal numbers for cycles modulo h and results about decomposing complete r-graphs into r-partite ones, and the new tools include the analysis of the shadow for extremal hypergraph problems and quantitative estimates for the bipartite canonical Ramsey problem.

3.1 Proof of the main result

An *r-partite r-graph* H is an r-graph with vertex set $\bigsqcup_{i=1}^{r} V_i$ (the V_is are pairwise disjoint), and every $e \in H$ satisfies $|e \cap V_i| = 1$ for all $i \in [r]$. When all such edges e are present, H is called a *complete r-partite r-graph.* When $|V_i| = s$ for all $i \in [r]$, a complete r-partite r-graph H is said to be *balanced*, and denoted $K_{s:r}$. For each integer $k \geq 1$, a *(loose, or linear) path of length k* denoted by P_k, is a collection of k edges $e_1, e_2, \ldots, e_k$ such that $|e_i \cap e_j| = 1$ if $i = j + 1$, and $e_i \cap e_j = \emptyset$ otherwise. We will often omit floors and ceilings in our calculations for ease of notation and all logs will have base 2.

We begin with a sketch of the proof of Theorem 3.0.2. The first step is to partition each of the hypergraphs that we must count into a bounded number of sub-hypergraphs, each of which can be encoded by an edge-colored graph (by choosing one pair from each hyperedge,

---

[1] *Recently, Conjecture 3.0.3 was proved by Balogh, Narayanan and Skokan (73) using the container method and Ferber, McKinley and Samotij (74) proved a more general result.*

and coloring that with the third vertex, in such a way that each pair is chosen at most once). This follows easily from a (straightforward) lemma of Kostochka, Mubayi and Verstraëte (60) which states that a 3-graph with no loose cycle of a given length contains a pair of vertices of bounded codegree.

The main task is therefore to bound the number of 3-graphs with no loose cycle of length $2l$ that can be encoded by an edge-colored graph G. Our strategy is to first partition the edges of G into complete bipartite graphs $K_{s_1,s_1}, \ldots, K_{s_m,s_m}$, with each $s_i \leq \log n$, such that $\sum_{i=1}^{m} s_i = O(n^2/\log n)$; this can be done greedily, using the Kővári–Sós–Turán Theorem (50); there are at most $2^{o(n^2)}$ choices for the sequence $(m, s_1, \ldots, s_m)$.

The problem is therefore reduced to counting the number of edge-colorings of a complete bipartite graph $K_{s,s}$ such that the associated 3-graph contains no loose cycle of length $2l$. Theorem 3.1.2 proves that there are at most $2^{O(s^2)} n^{O(s)}$ such colorings. This suffices to prove our main theorem, since we obtain the following upper bound for the number of 3-graphs with no loose cycle of length $2l$:

$$\begin{aligned}\left(\sum_{m,s_1,\ldots,s_m} \prod_{i=1}^{m} 2^{s_i^2} n^{s_i} \right)^{O(1)} &= \left(2^{o(n^2)} \cdot 2^{\sum_i s_i^2 + \sum_i s_i \log n} \right)^{O(1)} \\ &= \left(2^{o(n^2)} 2^{n^2 + O((n^2/\log n)\cdot \log n)} \right)^{O(1)} = 2^{O(n^2)}.\end{aligned}$$

3.1.1 Main technical statement

Given a graph G with $V(G) \subset [n]$, a *coloring function* is a function $\chi : G \to [n]$ such that $\chi(e) = z_e \in [n] \setminus e$ for every $e \in G$. We call z_e the *color* of e. The vector of colors $N_G = (z_e)_{e \in G}$ is called an *edge-coloring* of G. The pair (G, N_G) is an *edge-colored* graph. A *color class* is the set of all edges that receive the same color.

Given G, each edge-coloring N_G defines a 3-graph $H(N_G) = \{e \cup \{z_e\} : e \in G\}$, called the *extension of* G *by* N_G. When there is only one coloring that has been defined, we also use the notation $G^* = H(N_G)$ for the extension. Observe that any subgraph $G' \subset G$ also admits an extension by N_G, namely, $G'^* = \{e \cup \{z_e\} : e \in G'\} \subset G^*$. If $G' \subset G$ and $\chi|_{G'}$ is one-to-one, then G' is called *rainbow colored.* If a rainbow colored G' further satisfies that $z_e \notin V(G')$ for all $e \in G'$, then G' is said to be *strongly rainbow colored.* Note that a strongly rainbow colored graph $C_k \subset G'$ gives rise to 3-graph C_k in $G'^* \subset G^*$.

Definition 3.1.1. *For $k \geq 3, s \geq 1$, let $f(n, k, s)$ be the number of edge-colored complete bipartite graphs $G = K_{s,s}$ with $V(G) \subset [n]$, whose extension G^* is C_k-free.*

The function $f(n, k, s)$ allows us to encode 3-graphs, and our main technical theorem gives an upper bound for this function.

Theorem 3.1.2. *Fix an even integer $k \geq 4$. Then*

$$f(n, k, s) = 2^{O(s \log n + s^2)}$$

Note that the trivial upper bound is $f(n,k,s) \le n^{2s+s^2} \sim 2^{s^2 \log n}$ (first choose $2s$ vertices, then color each of its s^2 edges using an arbitrary vertex from $[n]$). The proof of Theorem 3.1.2 will be given in Sections 3–6.

3.1.2 A non-bipartite version of Theorem 3.1.2

Chung-Erdős-Spencer (75) and Bublitz (76) proved that the complete graph K_n can be decomposed into balanced complete bipartite graphs such that the sum of the sizes of the vertex sets in these bipartite graphs is at most $O(n^2/\log n)$. See also (77; 78) for some generalizations and algorithmic consequences. We state the result without proof as follows.

Theorem 3.1.3. *Let $n \ge 2$. Then, each n-vertex graph can be decomposed into complete bipartite graphs $K_{s_i,s_i}, i = 1, \ldots, m$, with $s_i \le \log n$ and $\sum_{i=1}^{m} s_i = O(n^2/\log n)$.*

Theorem 3.1.2 is about the number of ways to edge-color complete bipartite graphs with parts of size s and vertex set in $[n]$. Next, we use Theorems 3.1.2 and 3.1.3 to prove a related statement where we do not require the bipartite condition and the restriction to s vertices.

Definition 3.1.4. *For $k \ge 4$ and even, let $g(n,k)$ be the number of edge-colored graphs G with $V(G) \subset [n]$ such that the extension G^* is C_k-free.*

Lemma 3.1.5. *For fixed $k \ge 4$ and even,*

$$g(n,k) = 2^{O(n^2)}.$$

Proof. Given graph G, by applying Theorem 3.1.3, we may decompose G into balanced complete bipartite graphs $K_{s_1,s_1}, \dots, K_{s_m,s_m}$, with $s_i \le \log n$ and $\sum_{i=1}^m s_i = O(n^2/\log n)$. Then we trivially deduce the following two facts.

- From the second inequality, we have $m = O(n^2/\log n)$.
- Using the fact that these copies of $K_{s_i:s_i}$ are edge disjoint, we have

$$\sum_{i=1}^m s_i^2 \le \binom{n}{2} < n^2.$$

Therefore, to construct an edge-colored G, we need to first choose a sequence of positive integers $(m, s_1, \dots, s_m)$ such that $m \le c_1 n^2/\log n$, with some fixed $c_1 > 0$ and $s_i \le \log n$ for all i. More formally, let

$$S_n = \{(m, s_1, s_2, \dots, s_m) : m \le c_1 n^2/\log n, 1 \le s_i \le \log n, 1 \le i \le m\}.$$

Then

$$|S_n| \le \frac{c_1 n^2}{\log n} (\log n)^{\frac{c_1 n^2}{\log n}} = 2^{\log\left(\frac{c_1 n^2}{\log n}\right) + \frac{c_1 n^2 \log(\log n)}{\log n}} \le 2^{o(n^2)}. \tag{3.1}$$

Next, we sequentially construct an edge-colored K_{s_i,s_i} for each $i \in [m]$. Since G^* is C_k-free, $K^*_{s_i,s_i}$ is C_k-free. Writing v for a vector $(m, s_1, \dots, s_m)$ and applying Theorem 3.1.2 yields

$$g(n,k) \le \sum_{v \in S_n} \prod_{i=1}^m f(n,k,s_i) \le \sum_{v \in S_n} \prod_{i=1}^m 2^{O(s_i \log n + s_i^2)} \le \sum_{v \in S_n} 2^{O(\sum_{i=1}^m s_i \log n + s_i^2)}.$$

By Theorem 3.1.3 and Equation 3.1, this is at most

$$\sum_{v \in S_n} 2^{O((n^2/\log n)\cdot \log n + n^2)} = 2^{O(n^2)},$$

and the proof is complete. □

3.1.3 Proof of Theorem 3.0.2

A crucial statement that we use in our proof is that any r-graph such that every sub-edge has high codegree contains rich structures, including cycles. This was explicitly proved in (60) and we cite their following result.

Lemma 3.1.6. ***(Lemma 3.2 in (60))*** *For $r, k \geq 3$, if all sub-edges of an r-graph H have codegree greater than rk, then $C_k \subset H$.*

Now we have all the ingredients to complete the proof of our main result.

Proof of Theorem 3.0.2. Starting with any 3-graph H on $[n]$ with $C_k \not\subset H$, we claim that there exists a sub-edge with codegree at most $3k$. Indeed, otherwise all sub-edges of H will have codegree more than $3k$, and then by Lemma 3.1.6 we obtain a $C_k \subset H$. Let e' be the sub-edge of H with $0 < d_H(e') \leq 3k$ such that it has smallest lexicographic order among all such sub-edges. Delete all edges of H containing e' from H (i.e. delete $\{e \in H : e' \subset e\}$). Repeat this process of "searching and deleting" in the remaining 3-graph until there are no such sub-edges. We claim that the remaining 3-graph must have no edges at all. Indeed, otherwise we get a nonempty subgraph all of whose sub-edges have codegree greater than $3k$, and again by Lemma 3.1.6, we obtain a $C_k \subset H$.

Given any C_k-free 3-graph H on $[n]$, the algorithm above sequentially decomposes H into a collection of sets of at most $3k$ edges who share a sub-edge (a pair of two vertices) in common. We regard the collection of these pairs as a graph G. Moreover, for each edge $e \in G$, let N_e be the set of vertices $v \in V(H)$ such that $e \cup \{v\}$ is an edge of H at the time e was chosen. So $N_e \in \binom{[n]\setminus e}{\leq 3k}$, for all $e \in G$. Thus, we get a map

$$\phi : \mathrm{Forb}_3(n, C_k) \longrightarrow \left\{ (G, N_G) : G \subset \binom{[n]}{2}, N_G = \left(N_e \in \binom{[n] \setminus e}{\leq 3k} : e \in G \right) \right\}.$$

We observe that ϕ is injective. Indeed,

$$\phi^{-1}((G, N_G)) = H(N_G) = \{e \cup \{z_e\} : e \in G, z_e \in N_e\},$$

therefore $|\mathrm{Forb}_3(n, C_k)| = |\phi(\mathrm{Forb}_3(n, C_k))|$. Let $P = \phi(\mathrm{Forb}_3(n, C_k))$ which is the set of all pairs (G, N_G) such that $H(N_G)$ is C_k-free. Next we describe our strategy for upper bounding $|P|$.

For each pair $(G, N_G) \in P$ and $e \in G$, we pick exactly one $z_e^1 \in N_e$. Thus we get a pair (G_1, N_{G_1}), where $G_1 = G$, and $N_{G_1} = (z_e^1 : e \in G_1)$. Then, delete z_e^1 from each N_e, let $G_2 = \{e \in G_1 : N_e \setminus \{z_e^1\} \neq \emptyset\}$ and pick $z_e^2 \in N_e \setminus \{z_e^1\}$ to get the pair (G_2, N_{G_2}). For $2 \leq i < 3k$, we repeat this process for G_i to obtain G_{i+1}. Since each N_{G_i} contains only singletons, the pair

(G_i, N_{G_i}) can be regarded as an edge-colored graph. Note that we may get some empty G_is. This gives us a map

$$\psi : P \longrightarrow \left\{(G_1, \ldots, G_{3k}) : G_i \subset \binom{[n]}{2} \text{ is edge-colored for all } i \in [3k]\right\}.$$

Moreover, it is almost trivial to observe that ψ is injective, since if $y \neq y'$, then either the underlying graphs of y and y' differ, or the graphs are the same but the color sets differ. In both cases one can easily see that $\psi(y) \neq \psi(y')$. Again, we let $Q = \psi(P)$.

Note that $k \geq 4$ and even, by Lemma 3.1.5, we have

$$|\mathrm{Forb}_3(n, C_k)| = |P| = |Q| \leq \prod_{i=1}^{3k} g(n,k) = \prod_{i=1}^{3k} 2^{O(n^2)} = 2^{O(n^2)}.$$

$\square$

3.2 Proof of Theorem 3.1.2

The rest of the paper is devoted to the proof of Theorem 3.1.2. For simplicity of presentation, we write $k = 2l$ where $l \geq 2$. We first state our two main lemmas about edge-coloring bipartite graphs then give a proof of Theorem 3.1.2.

Lemma 3.2.1. *Let $l \geq 2, s, t \geq 1$, $G = K_{s,t}$ be an edge-colored complete bipartite graph with $V(G) \subset [n]$ and $Z = \{z_e : e \in G\} \subset [n]$ be the set of all colors. If G contains no strongly rainbow colored C_{2l}, i.e. the 3-uniform extension G^* of G is C_{2l}-free, then $|Z| < 2l(s+t)$.*

Lemma 3.2.2. *Let $l \geq 2, s, t \geq 1$, $D = D_l = (4l)2^{(4l)^7}$, $G = K_{s,t}$ be a complete bipartite graph with vertex set $V(G) \subset [n]$, and $Z \subset [n]$ be a fixed set of colors. Then the number of ways to edge-color G with Z such that the extension G^* contains no C_{2l}, is at most $D^{(s+t)^2}$.*

The proofs of these lemmas require several new ideas which will be presented in the rest of the paper. Here we quickly show that they imply Theorem 3.1.2.

Proof of Theorem 3.1.2. Recall that $l \geq 2$, and that $f(n, 2l, s)$ is the number of edge-colored copies of $K_{s,s}$ whose vertex set lies in $[n]$ and whose (3-uniform) extension is C_{2l}-free. To obtain such a copy of $K_{s,s}$, we first choose from $[n]$ its $2s$ vertices, then its at most $4ls$ colors by Lemma 3.2.1 and finally we color this $K_{s,s}$ by Lemma 3.2.2. This yields

$$f(n, 2l, s) \leq n^{2s+4ls} D^{(2s)^2} \leq 2^{5ls \log n + 4s^2 \log D} = 2^{O(s \log n + s^2)},$$

where the second inequality holds since $l \geq 2$. □

3.3 Proof of Lemma 3.2.1

In this section we prove Lemma 3.2.1. Our main tool is an extremal result about cycles modulo h in a graph. This problem has a long history, beginning with a Conjecture of Burr and Erdős that was solved by Bollobás (79) in 1976, see also (80; 81; 82; 83). In particular, we need the following lemma (see Diwan (81)) whose idea is based on considering the longest path in G and the neighbors of the two endpoints of the path.

Lemma 3.3.1. *If G is an n-vertex graph with at least $(h + 1)n$ edges, then G contains a cycle of length 2 modulo h.*

Recall that a rainbow colored cycle C_k is a copy of C_k with vertex set $V(C_k)$ in $[n]$ whose edges receives all distinct colors (where colors are vertices in $[n]$); whereas a strongly colored cycle C_k is rainbow colored and the set of all its colors is disjoint from its vertex set $V(C_k)$.

Lemma 3.3.2. *Let integers $l \geq 2$, $s, t \geq 1$, $G = K_{s,t}$ with $V(G) \subset [n]$ be edge-colored. If G contains a strongly rainbow colored cycle of length $2 \pmod{2l-2}$, then G contains a strongly rainbow colored C_{2l}.*

Proof. Let us assume that C is the shortest strongly rainbow colored cycle of length 2 modulo $2l-2$ in G. Then C has at least $2l$ edges. We claim that C is a C_{2l}. Suppose not, let e be a chord of C (such a chord exists as G is complete bipartite), such that C is cut up into two paths P_1 and P_2 by the two endpoints of e, and $|P_1| = 2l-1$. Let Z_1, Z_2 be the set of their colors respectively. If the color $z_e \notin Z_1 \cup V(P_1) \setminus e$, then $P_1 \cup e$ is a strongly rainbow colored cycle of length $2l$, a contradiction. Therefore $z_e \in Z_1 \cup V(P_1) \setminus e$, but then $z_e \notin Z_2 \cup V(P_2) \setminus e$, yielding a shorter strongly rainbow colored cycle $P_2 \cup e$ of length 2 modulo $2l-2$, a contradiction. $\square$

We now have all the necessary ingredients to prove Lemma 3.2.1.

Proof of Lemma 3.2.1. Suppose $|Z| \geq 2l(s+t)$. Then $|Z \setminus V(G)| \geq (2l-1)(s+t)$. For each color v in $Z \setminus V(G)$, pick an edge e of G with color v. We obtain a strongly rainbow colored subgraph G' of G with at least $(2l-1)(s+t)$ edges. Lemma 3.3.1 guarantees the existence of a rainbow colored cycle of length 2 modulo $2l-2$ in G'. By construction, this cycle is strongly rainbow. Lemma 3.3.2 then implies that there is a strongly rainbow colored C_{2l} in G. $\square$

3.4 Proof of Lemma 3.2.2

Our proof of Lemma 3.2.2 is inspired by the methods developed in (60). The main idea is to use the bipartite canonical Ramsey theorem. In order to use this approach we need to develop some new quantitative estimates for an asymmetric version of the bipartite canonical Ramsey theorem.

3.4.1 Canonical Ramsey theory

In this section we state and prove the main result in Ramsey theory that we will use to prove Lemma 3.2.2. We are interested in counting the number of edge-colorings of a bipartite graph, such that the (3-uniform) extension contains no copy of C_{2l}. The canonical Ramsey theorem allows us to find nice colored structures that are easier to work with. However, the quantitative aspects are important for our application and consequently we need to prove various bounds for bipartite canonical Ramsey numbers. We begin with some definitions.

Let G be a bipartite graph on vertex set with bipartition $X \sqcup Y$. For any subsets $X' \subset X$, $Y' \subset Y$, let $E_G(X', Y') = G[X' \sqcup Y'] = \{xy \in G : x \in X', y \in Y'\}$, and $e_G(X', Y') = |E_G(X', Y')|$. If X' contains a single vertex x, then $E_G(\{x\}, Y')$ will be simply written as $E_G(x, Y')$. The subscript G may be omitted if it is obvious from context.

Definition 3.4.1. *Let G be an edge-colored bipartite graph with $V(G) = X \sqcup Y$.*

- *G is* monochromatic *if all edges in $E(X, Y)$ are colored by the same color.*
- *G is* weakly X-canonical *if $E(x, Y)$ is monochromatic for each $x \in X$.*

- *G is X-*canonical *if it is* weakly X-canonical *and for all distinct $x, x' \in X$ the colors used on $E(x, Y)$ and $E(x', Y)$ are all different.*

In all these cases, the color z_x of the edges in $E(x, Y)$ is called a canonical color.

Lemma 3.4.2. *Let $G = K_{a,b}$ be an edge-colored complete bipartite graph with bipartition $A \sqcup B$, with $|A| = a, |B| = b$. If G is weakly A-canonical, then there exists a subset $A' \subset A$ with $|A'| = \sqrt{a}$ such that $G[A' \sqcup B] = K_{\sqrt{a},b}$ is A'-canonical or monochromatic.*

Proof. Take a maximal subset A' of A such that the coloring on $E(A', B)$ is A'-canonical. If $|A'| \geq \sqrt{a}$, then we are done. So, we may assume that $|A'| < \sqrt{a}$. By maximality of A', there are less then $\sqrt{a}$ canonical colors. By the pigeonhole principle, there are at least $|A|/|A'| \geq a/\sqrt{a} = \sqrt{a}$ vertices of A sharing the same canonical color, which gives a monochromatic $K_{\sqrt{a},b}$. □

Our next lemma guarantees that in an "almost" rainbow colored complete bipartite graph, there exists a rainbow complete bipartite graph.

Lemma 3.4.3. *For any integer $c \geq 2$, and $p > c^4$, if $G = K_{p,p}$ is an edge-colored complete bipartite graph, in which each color class is a matching, then G contains a rainbow colored $K_{c,c}$.*

Proof. Let $A \sqcup B$ be the vertex set of G. Pick two c-sets X, Y from A and B respectively at random with uniform probability. For any pair of monochromatic edges e, e', the probability that they both appear in the induced subgraph $E(X, Y)$ is

$$\left(\frac{\binom{p-2}{c-2}}{\binom{p}{c}} \right)^2 = \left(\frac{c(c-1)}{p(p-1)} \right)^2.$$

On the other hand, the total number of pairs of monochromatic edges is at most $p^3/2$, since every color class is a matching. Therefore the union bound shows that, when $p > c^4$, the probability that there exists a monochromatic pair of edges in $E(X, Y)$ is at most

$$\frac{p^3}{2}\left(\frac{c(c-1)}{p(p-1)}\right)^2 < \frac{pc^4}{2(p-1)^2} < 1.$$

Consequently, there exists a choice of X and Y such that the $E(X, Y)$ contains no pair of monochromatic edges. Such an $E(X, Y)$ is a rainbow colored $K_{c,c}$. $\square$

Now we are ready to prove the main result of this section which is a quantitative version of a result from (61). Note that the edge-coloring in this result uses an arbitrary set of colors. Since the conclusion is about "rainbow" instead of "strongly rainbow", it is not essential to have the set of colors disjoint from the vertex set of the graph.

Theorem 3.4.4 (Asymmetric bipartite canonical Ramsey theorem)**.** *For any integer $l \geq 2$, there exists real numbers $\epsilon = \epsilon(l) > 0, s_0 = s_0(l) = 2^{(4l)^7}$, such that if $G = K_{s,t}$ is an edge-colored complete bipartite graph on vertex set $X \sqcup Y$ with $|X| = s, |Y| = t$ with $s > s_0$ and $s/\log s < t \leq s$, then one of the following holds:*

- *G contains a rainbow colored $K_{4l,4l}$,*
- *G contains a $K_{q,2l}$ on vertex set $Q \sqcup R$, with $|Q| = q, |R| = 2l$ that is Q-canonical, where $q = s^{\epsilon}$,*
- *G contains a monochromatic $K_{q,2l}$ on vertex set $Q \sqcup R$, with $|Q| = q, |R| = 2l$, where $q = s^{\epsilon}$.*

Note that in the last two cases, it could be $Q \subset X, R \subset Y$ or the other way around.

Proof. We will show that $\epsilon = 1/18l$. First, fix a subset Y' of Y with $|Y'| = t^{1/4l}$ and let

$$W = \left\{ x \in X : \text{there exists a } Y'' \in \binom{Y'}{2l} \text{ such that } E_G(x, Y'') \text{ is monochromatic} \right\}.$$

If $|W| > s/2l$, then the number of $Y'' \in \binom{Y'}{2l}$ such that $E_G(x, Y'')$ is monochromatic for some x (with repetition) is greater than $s/2l$. On the other hand, $|\binom{Y'}{2l}| < |Y'|^{2l} = \sqrt{t}$. By the pigeonhole principle, there exists a $Y'' \in \binom{Y'}{2l}$ such that at least

$$\frac{s}{2l\sqrt{t}} \geq \frac{s}{2l\sqrt{s}} \geq s^{1/3}$$

vertices x have the property that $E_G(x, Y'')$ is monochromatic. Let Q_1 be a set of $s^{1/3}$ such x. Then we obtain a weakly Q_1-canonical $K_{s^{1/3},2l}$ on $Q_1 \sqcup Y''$ which, by Lemma 3.4.2, contains a canonical or monochromatic $K_{s^{1/6},2l}$. Since $\epsilon < 1/6$, this contains a $K_{s^\epsilon,2l}$ as desired.

We may now assume that $|W| \leq s/2l$. By definition of W and the pigeonhole principle, $E_G(x, Y')$ contains at least $|Y'|/2l$ (distinct) colors for every $x \in X \setminus W$. Hence, for each $x \in X \setminus W$ we can take $|Y'|/2l$ distinctly colored edges from $E(x, Y')$ to obtain a subgraph G' of G on $(X \setminus W) \sqcup Y'$ with $|X \setminus W||Y'|/2l$ edges.

Pick a subset $X' \subset X \setminus W$ with $|X'| = s^{1/16l^2}$ and $e_{G'}(X', Y') \geq |X'||Y'|/2l$. This is possible by an easy averaging argument. Let

$$Z = \left\{ y \in Y' : \text{there exists an } X'' \in \binom{X'}{2l} \text{ such that } E_{G'}(X'', y) \text{ is monochromatic} \right\}.$$

If $|Z| > |Y'|/20l$, then the number of $X'' \in \binom{X'}{2l}$ such that $E_{G'}(X'', y)$ is monochromatic for some y (with repetition) is greater than $|Y'|/20l$. On the other hand, $|\binom{X'}{2l}| < |X'|^{2l} = s^{1/8l}$. By the pigeonhole principle, there exists a $X'' \in \binom{X'}{2l}$ such that at least

$$\frac{|Y'|}{20ls^{1/8l}} = \frac{t^{1/4l}}{20ls^{1/8l}} \geq \frac{s^{1/4l}}{(\log s)^{1/4l} 20ls^{1/8l}} \geq s^{1/9l} = s^{2\epsilon}$$

vertices y have the property that $E_{G'}(X'', y)$ is monochromatic. Let Q_2 be a set of $s^{2\epsilon}$ such y. We find a weakly Q_2-canonical $K_{2l,s^{2\epsilon}}$ on $X'' \sqcup Q_2$. Again, by Lemma 3.4.2, a copy of $K_{2l,s^{\epsilon}}$ that is monochromatic or canonical is obtained.

Finally, we may assume that $|Z| \leq |Y'|/20l$. Then

$$\begin{aligned} e_{G'}(X', Y' \setminus Z) &\geq e_{G'}(X', Y') - |X'||Z| \geq \frac{1}{2l}|X'||Y'| - \frac{1}{20l}|X'||Y'| = \frac{9}{20l}|X'||Y'| \\ &\geq \frac{9}{20l}|X'||Y' \setminus Z|. \end{aligned}$$

Since each vertex $y \in Y' \setminus Z$ has the property that $E_{G'}(X', y)$ sees each color at most $2l-1$ times, for each $y \in Y' \setminus Z$ we may remove all edges from $E_{G'}(X', y)$ with duplicated colors (keep one for each color). We end up getting a bipartite graph G'' on $X' \sqcup (Y' \setminus Z)$ with at least $9|X'||Y' \setminus Z|/40l^2$ edges. By the Kővári–Sós–Turán theorem (50), there is a $c = c(l) > 0$ such that G'' contains a copy K of $K_{p,p}$ where $p > c \log s$. More precisely, writing $|X'| = m = s^{1/16l^2}$ and $|Y' \setminus Z| = n \geq (1 - 1/20l)t^{1/4l}$, the graph G'' contains a copy of $K_{p,p}$ if $|G''| \geq 2nm^{1-1/p} > (p-1)^{1/p} nm^{1-1/p} + (p-1)m$, which is an upper bound for the bipartite

Turán number for $K_{p,p}$. Since we have proved $|G''| > 9mn/40l^2$, a short calculation shows that we can let $p = (4l)^{-3} \log s$ and therefore $c(l) = 1/(4l)^3$.

Let $K = K_{p,p} \subset G''$ and $V(K) = A \sqcup B$. For each $x \in A$, the edge set $E(x, B)$ is rainbow colored, and for each $y \in B$, the edge set $E(A, y)$ is rainbow colored. Therefore each color class in K is a matching. By Lemma 3.4.3 and $s > s_0 = 2^{(4l)^4/c} = 2^{(4l)^7}$, we can find a rainbow colored $K_{4l,4l}$ in K as desired. $\square$

3.4.2 The induction argument for Lemma 3.2.2

We are now ready to prove Lemma 3.2.2. Let us recall the statement.

Lemma 3.2.2 *Let $l \geq 2, s, t \geq 1$, $D = D_l = (4l)2^{(4l)^7}$, $G = K_{s,t}$ be a complete bipartite graph with vertex set $V(G) \subset [n]$, and $Z \subset [n]$ be a fixed set of colors. Then the number of ways to edge-color G with Z such that the extension G^* contains no C_{2l}, is at most $D^{(s+t)^2}$.*

Here is a sketch of the proof. We proceed by induction on $s + t$. For the base cases $s + t = O(1)$ and $t < s/\log s$, we just upper bound the number of colorings using the trivial bound σ^{st}, where $\sigma = |Z|$ is less than $2l(s + t)$ by Lemma 3.2.1. For the induction step, we apply Theorem 3.4.4 to show that any coloring of $G = K_{s,t}$ contains a rainbow $K_{4l,4l}$ or a $K_{q,2l}$ that is either Q-canonical or monochromatic, where $q = |Q| = s^\epsilon$. The case of a rainbow $K_{4l,4l}$ is very easy to handle so we focus on the other two cases. So we are counting colorings of G that can be constructed as follows: first pick a q-set Q from X and a $2l$-set R from Y; next color $E(Q, R)$ in a Q-canonical or monochromatic fashion, then color $E(Q, Y \setminus R)$ to obtain a coloring of $E(Q, Y)$; finally color $E(X \setminus Q, Y)$.

In both cases (monochromatic and canonical), the number of ways to pick the q-set is at most s^q. The main step in the proof is to show that the number of ways to color $E(Q, Y)$ is bounded by $2^{O(qs)} = 2^{s^{1+\epsilon}}$, instead of the trivial $\sigma^{qs} = 2^{O(s^{1+\epsilon} \log s)}$. We will show this claim in the last two subsections. The idea is to first color a strongly rainbow path starting in Q since this creates restrictions on the possible colorings of the remaining edges.

Finally, the number of colorings of $E(X \setminus Q, Y)$ is at most $D^{(s+t-q)^2}$ by the induction hypothesis. Altogether, the number of colorings of $E(X, Y)$ is at most

$$s^q \cdot 2^{O(sq)} \cdot D^{(s+t-q)^2} \leq D^{(s+t)^2}.$$

Proof of Lemma 3.2.2. Let the vertex set of G be $S \sqcup T$ with $|S| = s$ and $|T| = t$. We apply induction on $s + t$. By Lemma 3.2.1, $|Z| := \sigma < 2l(s + t)$. The number of ways to color G is at most σ^{st}. As long as $s + t \leq D/2l$, we have

$$\sigma^{st} \leq D^{st} \leq D^{(s+t)^2}$$

and this concludes the base case(s).

For the induction step, we may henceforth assume $s + t > D/2l$, and the statement holds for all smaller values of $s + t$. Let us also assume without loss of generality that $t \leq s$.

Next, we deal with the case $t \leq s/\log s$. Since $D = (4l)2^{(4l)^7} > 16l^2$, $s > (s+t)/2 > D/4l > 4l$ and the number of ways to color G is at most

$$\sigma^{st} \leq (2l(s+t))^{st} \leq 2^{\frac{s^2 \log(2l(s+t))}{\log s}} \leq 2^{\frac{s^2 \log(4ls)}{\log s}} \leq 2^{2s^2} \leq 2^{(s+t)^2 \log D} = D^{(s+t)^2}.$$

Therefore, for the rest of the proof we may assume $s/\log s < t \leq s$ and $s > D/4l = 2^{(4l)^7}$. Since s and t differ from each other by only a little, we would like to give the two partite sets (recall that $V(G) = S \sqcup T$) of G pseudonyms, so that we can discuss the appearance of some coloring pattern of some subgraph and its symmetric case at the same time.

Definition 3.4.5. *Let $\{X, Y\} = \{S, T\}$. Given $X' \subset X$ and $Y' \subset Y$, let $\#E(X', Y')$ be the number of ways to color the edges in $E(X', Y')$.*

The following lemma provides the essential idea of the induction step.

Claim 3.4.6. *Let $q = s^\epsilon < s/\log s < t < s$ and $s > D/4l = 2^{(4l)^7}$. Let $G = K_{s,t}$ on the vertex set $S \sqcup T$, let $\{X, Y\} = \{S, T\}$. Suppose that there is a subset $Q \subset X$ with $|Q| = q$ such that $\#E(Q, Y) \leq 2^{70l^2qs}$. Then $\#E(X, Y) < D^{(s+t)^2}/(4s^q)$.*

Proof. Delete Q from X and apply the induction hypothesis (of Lemma 3.2.2) to obtain $\#E(X \setminus Q, Y) \leq D^{(s+t-q)^2}$. Together with the condition $\#E(Q, Y) \leq 2^{70l^2qs}$ we have

$$s^q \cdot \#E(X, Y) \leq s^q \cdot \#E(Q, Y) \cdot \#E(X \setminus Q, Y).$$

Take logs, we get

$$
\begin{aligned}
q\log s + \log \#E(X,Y) &\le q\log s + 70l^2qs + \log\left(D^{(s+t-q)^2}\right)\\
&= q\log s + 70l^2qs + \left(q^2 - 2q(s+t) + (s+t)^2\right)\log D\\
&= q\log s + 70l^2qs - 2qs\log D + q(q-2t)\log D + (s+t)^2\log D\\
&\le 71l^2qs - 2qs\log D + (s+t)^2\log D\\
&= (71l^2 - 2\log D)qs + (s+t)^2\log D\\
&< \log\frac{1}{4} + (s+t)^2\log D,
\end{aligned}
$$

where the second inequality holds since $\log s < s$ and $q - 2t < 0$, while the last inequality holds since we take $D = (4l)2^{(4l)^7}$, so $(71l^2 - 2\log D)qs < -2$. Therefore, we have $\#E(X,Y) < D^{(s+t)^2}/(4s^q)$. □

Since we may assume that $s/\log s < t \le s$, and $s > D/4l = 2^{(4l)^7} = s_0$, the conditions of Theorem 3.4.4 hold. Let $N_G = (z_e)_{e\in G}$ be an edge-coloring of G using colors in Z. By Theorem 3.4.4, such an edge-colored G will contain a subgraph G' that is either

- a rainbow colored $K_{4l,4l}$, or
- a Q-canonical $K_{q,2l}$, or
- a monochromatic $K_{q,2l}$,

where $|Q| = q = s^\epsilon$ and $\epsilon = 1/18l$.

Claim 3.4.7. *G' cannot be a rainbow colored $K_{4l,4l}$.*

Proof of Claim 3.4.7. Suppose for a contradiction that $G' = K_{4l,4l}$ is rainbow colored and Z' is the set of colors used on G'. Then $|Z' \backslash V(G')| \geq 16l^2 - 8l$. Pick an edge of each color in $Z' \backslash V(G')$ to obtain a strongly rainbow colored subgraph G'' of G' with $|G''| = 16l^2 - 8l \geq (2l-1)8l$. By Lemma 3.3.1, G'' contains a strongly rainbow colored cycle of length $2 \mod 2l-2$. Lemma 3.3.2 now implies the existence of a strongly rainbow colored C_{2l} in G'', which forms a linear C_{2l} in G^*, a contradiction. $\square$

Thus, we are guaranteed that for each edge-coloring of G that we want to count there is a subgraph $G' = K_{q,2l}$ of G that is colored in either a Q-canonical or monochromatic fashion. Let the vertex set of G' be $Q \sqcup R$, where $Q \in \binom{X}{q}$ and $R \in \binom{Y}{2l}$. Define $|X| = a, |Y| = b$, so $\{a, b\} = \{s, t\}$.

There are four combinations according to the choice of (X, Y) and the coloring patterns. If we can show that in each case $\#E(Q, Y) \leq 2^{70l^2qs}$, then we are done. Because in each case, to count the number of colorings, we may first choose a q-set from X then apply Claim 3.4.6 to color G. The total number of colorings is at most

$$s^q \cdot \#E(X, Y) \leq s^q \cdot \frac{D^{(s+t)^2}}{4s^q} = \frac{D^{(s+t)^2}}{4}.$$

Therefore, our goal of the last two subsections is to show the following. When $Q \subset X$ is fixed, if there exists an $R \subset Y$ such that $E(Q, R)$ is either Q-canonical or monochromatic, we have $\#E(Q, Y) \leq 2^{70l^2qs}$.

3.4.2.1 The canonical case

Recall that for each $x \in Q$, the edges in $E(x, R)$ all have the same color z_x which is called a *canonical color*. Let $Z_c = \{z_x : x \in Q\}$ be the set of all canonical colors. For each edge xy with $x \in Q, y \in Y \setminus (R \cup Z_c)$, a color $z_{xy} \neq z_x$ is called a *free color*. We will count the number of colorings of $E(Q, Y)$, and then remove Q to apply the induction hypothesis. For each coloring N_G, consider the following partition of $Y \setminus (R \cup Z_c)$ into two parts:

$$Y_0 = \{y \in Y \setminus (R \cup Z_c) : E(y, Q) \text{ sees at most } 11l - 1 \text{ distinct free colors}\},$$

$$Y_1 = \{y \in Y \setminus (R \cup Z_c) : E(y, Q) \text{ sees at least } 11l \text{ distinct free colors}\}.$$

We claim that the length of strongly rainbow colored paths that lie between Q and Y_1 is bounded.

Claim 3.4.8. *If there exists a strongly rainbow colored path $P = P_{2l-2} \subset E(Q, Y_1)$ with both end-vertices $u, v \in Q$, then there exists a C_{2l} in G^*.*

Proof of Claim 3.4.8. Clearly, P extends to a linear P_{2l-2} in G^*. We may assume both $z_u, z_v \notin V(P^*)$, where $P^* = \{e \cup \{z_e\} : e \in P\}$ is the extension of P. Otherwise, suppose w.l.o.g. $z_u \in V(P^*)$, let y be the vertex next to u in P, let S_y be of maximum size among sets

$$\{x \in Q : xy \text{ all colored by distinct free colors}\}.$$

Since $y \in Y_1$, $|S_y| \geq 11l$. Note that $|V(P^*)| = 4l - 3$ and $|V(P^*) \cap Y_1| \geq l - 1$, we have $|S_y \setminus V(P^*)| \geq 11l - (4l - 3 - (l-1)) \geq 8l$. Since $|V(P^*)| < 4l$, $E(y, S_y)$ is rainbow, and G' is Q-canonical, there must be at least $4l$ vertices in $S_y \setminus V(P^*)$ whose canonical color is not in $V(P^*)$. Among these $4l$ vertices there is at least one u' with $z_{u'y} \notin V(P^*)$. Replacing u by u', we get a strongly rainbow colored path of length $2l - 2$ with $z_u \notin V(P^*)$.

Now, Since $|R| = 2l$, we can find a vertex $y \in R$ such that $y \notin \{z_e : e \in P\}$. Further, since both $z_u, z_v \notin V(P^*)$ and $z_u \neq z_v$, the set of edges

$$P^* \cup \{uyz_u, vyz_v\}$$

forms a copy of C_{2l} in G^*. □

Thanks to this observation about strongly rainbow paths, we can bound the number of colorings on $E(Q, Y_1)$ as follows. It is convenient to use the following notation.

Claim 3.4.9. $\#E(Q, Y_1) \leq (2l)^q \cdot (32l^2)^{bq} \cdot (qb)^{2lq} \cdot \sigma^{6lq+8l^2b}$.

Proof of Claim 3.4.9. By Claim 3.4.8, according to the length of the longest strongly rainbow colored path starting at a vertex, Q can be partitioned into $2l - 3$ parts $\bigsqcup_{i=1}^{2l-3} Q_i$, where

$$\begin{aligned} Q_i = \{x \in Q : &\text{ the longest strongly rainbow colored path} \\ &\text{starting at } x \text{ and contained in } E(Q, Y_1) \text{ has length } i\}. \end{aligned}$$

For each i, let $q_i = |Q_i|$. We now bound the number of colorings of the edges in $E(Q_i, Y_1)$.

Firstly, for each $x \in Q_i$, choose an i-path $P_x \subset E(Q, Y_1)$ starting at x and color it strongly rainbow. The number of ways to choose and color these paths for all the vertices $x \in Q_i$ is

$$\prod_{x \in Q_i} \#P_x \leq \left((qb)^{\lceil (i+1)/2 \rceil} \sigma^i \right)^{q_i} \leq (qb\sigma)^{iq_i}.$$

Fix an $x \in Q_i$. Partition Y_1 into 3 parts depending on whether y is on the extension P_x^* of the path starting at x, or the color of xy is on P_x^* or else, i.e. $Y_1 = \bigsqcup_{j=1}^{3} Y_{i,x}^{(j)}$, where

$$\begin{aligned}
Y_{i,x}^{(1)} &= Y_1 \cap V(P_x^*), \\
Y_{i,x}^{(2)} &= \{y \in Y_1 \setminus Y_{i,x}^{(1)} : z_{xy} \in V(P_x^*)\}, \\
Y_{i,x}^{(3)} &= Y_1 \setminus (Y_{i,x}^{(1)} \cup Y_{i,x}^{(2)}).
\end{aligned}$$

Depending on the part of Y_1 that a vertex y lies in, we can get different restrictions on the coloring of the edges in $E(y, Q_i)$.

- If $y \in Y_{i,x}^{(1)}$, then z_{xy} has as many as σ choices. Note that $|P_x^*| = 2i + 1$, and $|Y_{i,x}^{(1)}| \leq i + \lceil i/2 \rceil \leq 2i$. This gives $\#E(x, Y_{i,x}^{(1)}) \leq \sigma^{2i}$.
- If $y \in Y_{i,x}^{(2)}$, then $z_{xy} \in V(P_x^*)$, so there are at most $2i + 1$ choices for this color and $\#E(x, Y_{i,x}^{(2)}) \leq (2i + 1)^b$.
- Lastly, let $|Y_{i,x}^{(3)}| = b_{i,x}$. If $y \in Y_{i,x}^{(3)}$, then xy extends P_x into a strongly rainbow colored path $P'_x = P_x \cup \{xy\}$ of length $i+1$, which forces the edges $x'y$ to be colored by $V(P_x'^*)$ for each $x' \in Q_i \setminus V(P_x'^*)$. Otherwise, the path $P'_x \cup \{x'y\}$ is a strongly rainbow colored path

of length $i+2$ starting at a vertex $x' \in Q_i$, contradicting the definition of Q_i. Therefore, $z_{x'y}$ has at most $2i+3$ choices if $x' \in Q_i \setminus V(P_x'^*)$. Putting this together, for each $y \in Y_{i,x}^{(3)}$, we have

$$\#E(Q_i \setminus V(P_x'^*), y) \le (2i+3)^{q_i}.$$

Noticing that $|Q_i \cap V(P_x'^*)| \le i + 1 + \lceil (i+1)/2 \rceil \le 2i+1$, we have

$$\#E(Q_i, y) \le \#E(Q_i \cap V(P_x'^*), y) \cdot \#E(Q_i \setminus V(P_x'^*), y) \le \sigma^{2i+1}(2i+3)^{q_i}.$$

Counting over all $y \in Y_{i,x}^{(3)}$, we have

$$\#E(Q_i, Y_{i,x}^{(3)}) \le \prod_{y \in Y_{i,x}^{(3)}} \#E(Q_i, y) \le \sigma^{(2i+1)b_{i,x}}(2i+3)^{q_i b_{i,x}}.$$

Hence the number of ways to color $E(x, Y_1) \cup E(Q_i, Y_{i,x}^{(3)})$ is at most

$$2^b \cdot \#E(x, Y_{i,x}^{(1)}) \cdot \#E(x, Y_{i,x}^{(2)}) \cdot \#E(Q_i, Y_{i,x}^{(3)}) \le 2^b \cdot \sigma^{2i} \cdot (2i+1)^b \cdot \sigma^{(2i+1)b_{i,x}}(2i+3)^{q_i b_{i,x}}.$$

The term 2^b arises above since $Y_{i,x}^{(1)}$ has already been fixed by choosing and coloring P_x, so we just need to partition $Y_1 \setminus Y_{i,x}^{(1)}$ to get $Y_{i,x}^{(2)}$ and $Y_{i,x}^{(3)}$.

Now we remove x from Q_i, $Y^{(3)}_{i,x}$ from Y_1 and repeat the above steps until we have the entire $E(Q_i, Y_1)$ colored. Note that $\sum_{x \in Q_i} b_{i,x} \leq b$, and that $i \leq 2l-3$ which implies $2i+3 < 4l$. We obtain

$$\begin{aligned}
\#E(Q_i, Y_1) &\leq \prod_{x \in Q_i} \#P_x \cdot \# \left(E(x, Y_1) \cup E(Q_i, Y^{(3)}_{i,x}) \right) \\
&\leq (qb\sigma)^{iq_i} \prod_{x \in Q_i} 2^b \cdot \sigma^{2i} \cdot (2i+1)^b \cdot \sigma^{(2i+1)b_{i,x}} (2i+3)^{q_i b_{i,x}} \\
&\leq (qb\sigma)^{2lq_i} + \prod_{x \in Q_i} 2^b \cdot \sigma^{4l+4lb_{i,x}} \cdot (4l)^{b+q_i b_{i,x}} \\
&\leq (qb\sigma)^{2lq_i} \cdot 2^{bq_i} \cdot \sigma^{4lq_i+4lb} \cdot (4l)^{bq_i+bq_i} \\
&= (32l^2)^{bq_i} \cdot (qb)^{2lq_i} \cdot \sigma^{6lq_i+4lb}.
\end{aligned}$$

Because $\sum_{i=1}^{2l-3} q_i = q$, taking the product over $i \in [2l-3]$, we obtain

$$\begin{aligned}
\#E(Q, Y_1) \leq (2l-3)^q \prod_{i=1}^{2l-3} \#E(Q_i, Y_1) &\leq (2l-3)^q \prod_{i=1}^{2l-3} (32l^2)^{bq_i} \cdot (qb)^{2lq_i} \cdot \sigma^{6lq_i+4lb} \\
&\leq (2l)^q \cdot (32l^2)^{bq} \cdot (qb)^{2lq} \cdot \sigma^{6lq+8l^2 b},
\end{aligned}$$

where $(2l-3)^q$ counts the number of partitions of Q into the Q_i. □

Since $G' = E(Q, R)$ is Q-canonical,

$$\#E(Q, R) \leq \sigma^q.$$

As $|Z_c| \leq q$,

$$\#E(Q, Y \cap Z_c) \leq \sigma^{q^2}.$$

By definition of Y_0,

$$\#E(Q, Y_0) \leq (\sigma^{11l}(11l+1)^q)^b \leq (\sigma^{11l}(12l)^q)^b.$$

Therefore to color $E(Q, Y)$, we need to first choose the subsets R and $Z_c \cap Y$ of Y and then take a partition to get Y_0 and Y_1. We color each of $E(Q, R), E(Q, Y \cap Z_c), E(Q, Y_0)$ and $E(Q, Y_1)$. This gives

$$\begin{aligned}
\#E(Q, Y) &\leq b^{2l} b^q 2^b \cdot \#E(Q, R) \cdot \#E(Q, Y \cap Z_c) \cdot \#E(Q, Y_0) \cdot \#E(Q, Y_1) \\
&\leq b^{2l} b^q 2^b \cdot \sigma^q \cdot \sigma^{q^2} \cdot (\sigma^{11l}(12l)^q)^b \cdot [(2l)^q \cdot (32l^2)^{bq} \cdot (qb)^{2lq} \cdot \sigma^{6lq+8l^2 b}] \\
&= b^{2l} 2^b \cdot (2lb)^q \cdot (384l^3)^{qb} \cdot (qb)^{2lq} \cdot \sigma^{q^2+(6l+1)q+(8l^2+11l)b}.
\end{aligned}$$

Recall that $q = s^\epsilon < s/\log s < t \leq s$, $\sigma \leq 2l(s+t) \leq 4ls$, and $s > 2^{(4l)^7} > 4l$. There are two cases according to the choices of a and b, i.e. $(a, b) = (s, t)$ and $(a, b) = (t, s)$. But in either case, we have $b \leq s$, hence

$$\begin{aligned}
\#E(Q, Y) &\leq s^{2l} 2^s \cdot (2ls)^q \cdot (384l^3)^{qs} \cdot (qs)^{2lq} \cdot (4ls)^{q^2+(6l+1)q+(8l^2+11l)s} \\
&\leq s^{2l} 2^s \cdot s^{2q} \cdot (384l^3)^{qs} \cdot s^{4lq} \cdot s^{2q^2+2(6l+1)q+2(8l^2+11l)s}.
\end{aligned}$$

Take logs,

$$\begin{aligned}
\log \#E(Q,Y) &\le \big(2l + 2q + 4lq + 2q^2 + 2(6l+1)q + 2(8l^2 + 11l)s\big)\log s + s + qs\log(384l^3)\\
&\le 3q^2\log s + (16l^2 + 22l + 1)s\log s + qs\left(\log 384 + \log(l^3)\right)\\
&\le l^2 s\log s + 27l^2 s\log s + qs\,(9 + 3\log l)\\
&= 28l^2 s\log s + 2l^2 qs\\
&\le 28l^2 qs + 2l^2 qs = 30l^2 qs < 70l^2 qs.
\end{aligned}$$

The last inequality holds since $\log s < s^{1/18l} = s^{\epsilon} = q$ when $s > 2^{(4l)^7}$.

3.4.2.2 The monochromatic case

Recall that the vertex set of $G' = K_{q,2l}$ is $Q \sqcup R$, where $Q \in \binom{X}{q}$ and $R \in \binom{Y}{2l}$. The term *canonical color* now refers to the only color z_c that is used to color all edges of G', and $Z_c = \{z_c\}$ still means the set of canonical colors. A *free color* is a color that is not z_c. As before we will count the number of colorings of $E(Q,Y)$, and then remove Q to apply the induction hypothesis.

Let $Y_1 = Y \setminus (R \cup Z_c)$. Similar to Claim 3.4.8, we claim that the length of a strongly rainbow colored path between Q and Y_1 is bounded.

Claim 3.4.10. *If there exists a strongly rainbow colored path $P = P_{4l-2} \subset E(Q,Y_1)$ with both end-vertices $u, v \in Q$, then there exists a C_{2l} in G^*.*

Proof of Claim 3.4.10. We observe that z_c appears in the path or the color of the path at most once, as P is strongly rainbow. Hence, by the pigeonhole principle, there exists a sub-path P' of length $2l-2$ such that $z_c \notin V(P'^*)$ and both end-vertices u, v of P' are in Q.

Now, Since $|R| = 2l$, we can find two vertices $y, y' \in R$ such that $y, y' \notin \{z_e : e \in P'\}$. Thus, the edges

$$P'^* \cup \{uyz_c, vy'z_c\}$$

yield a copy of C_{2l} in G^*. $\square$

Again, we first use this claim to color $E(Q, Y_1)$.

Claim 3.4.11. $\#E(Q, Y_1) \leq (4l)^q \cdot (128l^2)^{qb} \cdot (qb)^{4lq} \cdot \sigma^{12lq+32l^2b}$.

Proof of Claim 3.4.11. The proof proceeds exactly the same as that of Claim 3.4.9, except that Q is partitioned into $4l-3$ parts $\bigsqcup_{i=1}^{4l-3} Q_i$. So in the calculation at the end, we have $i \leq 4l-3$ which gives $2i+3 < 8l$ and

$$\begin{aligned}
\#E(Q_i, Y_1) &\leq \prod_{x \in Q_i} \#P_x \cdot \# \left(E(x, Y_1) \cup E(Q_i, Y_{i,x}^{(3)})\right) \\
&\leq (qb\sigma)^{iq_i} \prod_{x \in Q_i} 2^b \cdot \sigma^{2i} \cdot (2i+1)^b \cdot \sigma^{(2i+1)b_{i,x}} (2i+3)^{q_i b_{i,x}} \\
&\leq (qb\sigma)^{4lq_i} \prod_{x \in Q_i} 2^b \cdot \sigma^{8l+8lb_{i,x}} \cdot (8l)^{b+q_i b_{i,x}} \\
&\leq (qb\sigma)^{4lq_i} \cdot 2^{bq_i} \cdot \sigma^{8lq_i+8lb} \cdot (8l)^{bq_i+bq_i} \\
&\leq (128l^2)^{bq_i} \cdot (qb)^{4lq_i} \cdot \sigma^{12lq_i+8lb}.
\end{aligned}$$

Again, note that $\sum_{i=1}^{4l-3} q_i = q$. Taking the product over $i \in [4l-3]$, we obtain

$$\#E(Q, Y_1) \leq (4l-3)^q \prod_{i=1}^{4l-3} \#E(Q_i, Y_1) \leq (4l-3)^q \prod_{i=1}^{4l-3} (128l^2)^{bq_i} \cdot (qb)^{4lq_i} \cdot \sigma^{12lq_i+8lb}$$
$$\leq (4l)^q \cdot (128l^2)^{qb} \cdot (qb)^{4lq} \cdot \sigma^{12lq+32l^2b},$$

where $(4l-3)^q$ counts the number of partitions of Q into the Q_i. □

Similarly, to color $E(Q, Y)$, we need to choose the subsets R and $Y \cap Z_c$, and what remains is Y_1. Consequently,

$$\#E(Q, Y) \leq b^{2l}b \cdot \#E(Q, R) \cdot \#E(Q, Y \cap Z_c) \cdot \#E(Q, Y_1)$$
$$\leq b^{2l}b \cdot \sigma \cdot \sigma^q \cdot [(4l)^q \cdot (128l^2)^{qb} \cdot (qb)^{4lq} \cdot \sigma^{12lq+32l^2b}]$$
$$= b^{2l+1}(4l)^q \cdot (128l^2)^{qb} \cdot (qb)^{4lq} \cdot \sigma^{1+(12l+1)q+32l^2b}.$$

Again, recall that $q = s^\epsilon < s/\log s < t \leq s$, $\sigma \leq 2l(s+t) \leq 4ls$ and $s > 2^{(4l)^7} > 4l$. There are two cases according to the choices of a and b, i.e. $(a, b) = (s, t)$ and $(a, b) = (t, s)$. In either case $b \leq s$, hence

$$\#E(Q, Y) \leq s^{2l+1}(4l)^q \cdot (128l^2)^{qs} \cdot (qs)^{2lq} \cdot (4ls)^{1+(12l+1)q+32l^2s}$$
$$\leq s^{2l+1}(4l)^q \cdot (128l^2)^{qs} \cdot (s)^{4lq} \cdot s^{2+2(12l+1)q+64l^2s}$$

Take logs,

$$
\begin{aligned}
\log \#E(Q,Y) &\leq (2l + 1 + 4lq + 2 + 2(12l+1)q + 64l^2 s)\log s + q\log(4l) + qs\log(128l^2) \\
&\leq 65l^2 s\log s + qs\,(7 + 2\log l) \\
&\leq 70l^2 qs.
\end{aligned}
$$

Again, the last inequality holds since $\log s < q$ when $s > 2^{(4l)^7}$. □

3.5 Concluding remarks

• A straightforward corollary of Theorem 3.0.2 is the very same result for hypergraph paths P_k. Indeed, for the upper bound on $\mathrm{Forb}_r(n, P_k)$ one has to just observe that $P_k \subset C_{2\lceil (k+1)/2\rceil}$, while the lower bound is trivial.

• The main open problem raised by our work is to solve the analogous question for larger r and for odd cycles (Conjecture 3.0.3).

For $r = 3$, our method will not work for odd cycles as it relies on finding a bipartite structure from which it is difficult to extract odd 3-uniform cycles (although this technical hurdle could be overcome to solve the corresponding extremal problem in (60)).

For larger r, our method does not work because the cost of decomposing a complete r-graph into complete r-partite subgraphs is too large to remain an error term. More precisely, for $r = 3$, we implicitly applied Lemma 3.2.2 (in the proof of Lemma 3.1.5) to reduce the number of ways to color a graph to at most $2^{O(n^2)}$ instead of the trivial $2^{O(n^2 \log\log n)}$. But for $r > 3$ the main term in the calculation turns out to be $2^{O(n^{r-1}(\log n)^{(r-3)/(r-2)})}$ which comes from choosing the colors for the copies of $K_{s_i:r-1}$. This cannot be improved due to Theorem 3.1.3 and Turán type result of even cycles each of which gives a bound that is sharp in order of magnitude. Consequently, even if we proved a version of Lemma 3.2.2 for $r > 3$ (and the tools we have developed should suffice to provide such a proof) this would not improve Theorem 3.0.2 for $r > 3$.

- Another way to generalize the result of Morris-Saxton to hypergraphs is to consider similar enumeration questions when the underlying r-graph is linear, meaning that every two edges share at most one vertex. Here the extremal results have recently been proved in (84) and the formulas are similar to the case of graphs. The special case of this question for linear triple systems without a C_3 is related to the Ruzsa-Szeméredi $(6, 3)$ theorem and sets without 3-term arithmetic progressions.

CHAPTER 4

THE THRESHOLD FUNCTIONS IN EXTREMAL SET THEORY

4.1 Configurations formed by subsets of $[n]$

Fix $t \geq 2$ and a vector $b = (b_S)_{S \subset [t]}$ where $b_S \in \{0, 1\}$ for all $S \subset [t]$. Let $\mathcal{F}_b$ denote the family of all collections of subsets $F_1, \ldots, F_t$ of $[n]$ such that for all $S \subset [t]$

$$F_S \equiv \bigcap_{i \in S} F_i \bigcap_{i \in [t] \setminus S} F_i^c \neq \emptyset \quad \text{iff} \quad b_S = 1.$$

Clearly, $S \neq T$ implies that $F_S \cap F_T = \emptyset$ and then for each $S \subset [n]$

$$F_i = \bigcup_{i \in S} F_S. \tag{4.1}$$

In context, collections that belong to $\mathcal{F}_b$ are called *members of* $\mathcal{F}_b$ or $\mathcal{F}_b$*-configurations*. Let $\mathcal{H}_b$ be the hypergraph with vertex set $V(\mathcal{H}_b) = \{S \subset [t] : b_S = 1\}$ and edge set $\mathcal{H}_b = \{H_1, \ldots, H_t\}$, where $H_i = \{S : i \in S, b_S = 1\}$. Note that H_i could be the empty set. Let $v = |V(\mathcal{H}_b)|$ and $e = |\mathcal{H}_b|$. Note that $e \leq t$, and for the case that there are no repeated sets in the collection F_is, we have $e = t$.

For $\mathcal{G} \subset \mathcal{H}_b$ with $\mathcal{G} = \{G_1, \ldots, G_s\}$ $(1 \leq s \leq e)$, partition $V(\mathcal{H}_b)$ into *atoms* $A_1, \ldots, A_k$, where the A_j's are the smallest disjoint sets in the σ-algebra generated by $\mathcal{G}$ on $V(\mathcal{H}_b)$. Alternatively, to each $x \in V(\mathcal{H}_b)$, associate the vector $x_{\mathcal{G}} = (\chi_i(x))_{i=1}^{s}$, where χ_i is the characteristic function of the set G_i. Then $A_1, \ldots, A_k$ are the equivalence classes of vertices where $x \sim x'$ if and only if $x_{\mathcal{G}} = x'_{\mathcal{G}}$. Put $a_i = |A_i|$ for all i.

Example:

$$V(\mathcal{H}_b) = \{e_1 = \{1,2,3\}, e_2 = \{1,2,4\}, e_3 = \{2,3\}, e_4 = \{3,4\}, e_5 = \{4\}\}.$$

$v = 5, t = e = 4.$

$$H_1 = \{e_1, e_2\}, H_2 = \{e_1, e_2, e_3\}, H_3 = \{e_1, e_4\}, H_4 = \{e_4, e_5\}.$$

$\mathcal{G} = \{H_1, H_2\}.$

$$A_1 = \{e_1, e_2\}, A_2 = \{e_3\}, A_3 = \{e_4, e_5\}.$$

$$(e_1)_{\mathcal{G}} = (1,1), (e_2)_{\mathcal{G}} = (1,1), (e_3)_{\mathcal{G}} = (0,1), (e_4)_{\mathcal{G}} = (0,0), (e_5)_{\mathcal{G}} = (0,0).$$

Observe that for any member $\mathcal{F}$ of $\mathcal{F}_b$, the t sets of $\mathcal{F}$ form a Venn diagram on $[n]$ with 2^t cells, while only $v = |V(\mathcal{H}_b)|$ of the cells are non-empty. Each non-empty cell F_S is labeled by a unique set $S \subset [t]$ with $b_S = 1$. The members of $\mathcal{F}_b$ give rise uniquely to ordered (labeled) partitions of $[n]$ into v non-empty parts that we can write as $\{P_S : b_S = 1\}$. Each F_i in $\mathcal{F} = (F_1, F_2, \ldots, F_t) \in \mathcal{F}_b$ is then equal to $\bigcup_{S \in H_i} P_S$. Inclusion-exclusion gives us

$$|\mathcal{F}_b| = \sum_{i=0}^{v} (-1)^i \binom{v}{i} v^{n-i} = \Theta(v^n). \tag{4.2}$$

4.2 Threshold functions

For $p \in [0,1]$ and an integer n, let $Q(n,p)$ be the random set system, obtained by picking each subset of $[n]$ independently with probability p.

Since $Q(n,p)$ contains no repeated sets, when looking at configurations $\mathcal{F}_b$ that appear in it, we may assume that b is chosen such that members of $\mathcal{F}_b$ contain no repeated sets. Notice that $e = |\mathcal{H}_b|$ represents the number of distinct sets in each member of $\mathcal{F}_b$, so $t = e$.

Let P be a property of a realization of the set system $Q(n,p)$. For instance, the property that we are talking about primarily in this paper is that P="there is a member of $\mathcal{F}_b$ that appear as a sub-hypergraph of $Q(n,p)$". We define two types of threshold functions:

$p_0 = p_0(n)$ is called a *threshold function* for a property P if

- when $p = o(p_0), \lim_{n\to\infty} \mathbb{P}(Q(n,p) \models P) = 0$,
- when $p = \omega(p_0), \lim_{n\to\infty} \mathbb{P}(Q(n,p) \models P) = 1$,

or vice versa.

$p_0 = p_0(n)$ is called a *weak threshold function* for a property P if there exists an $r > 0$

- when $p = o(p_0/n^r), \lim_{n\to\infty} \mathbb{P}(Q(n,p) \models P) = 0$,
- when $p = \omega(p_0 n^r), \lim_{n\to\infty} \mathbb{P}(Q(n,p) \models P) = 1$,

or vice versa.

Obviously, $p = o(p_0/n^r) \Rightarrow p = o(p_0)$ and $p = \omega(p_0 n^r) \Rightarrow p = \omega(p_0)$, so a threshold function p_0 is also a weak threshold function, but not conversely.

4.3 Main theorems

4.3.1 Basic second moment

Our main results are using the second moment method to prove the existence of certain configurations in $Q(n,p)$ with some threshold functions. Recall that for $\mathcal{G} \subset \mathcal{H}_b$, $A_1, \ldots, A_k$ are the smallest disjoint sets in the σ-algebra generated by $\mathcal{G}$ on $V(\mathcal{H}_b)$. Let $|A_i| = a_i$ for all $i \in [k]$.

Theorem 4.3.1 (Basic Second Moment)**.** *Fix $t \geq 2$ and a 0-1 vector $b = (b_S)_{S \subset [t]}$. Suppose that for all $\mathcal{G} \subset \mathcal{H}_b$ with $|\mathcal{G}| = s$ and $1 \leq s \leq e$,*

$$\sum_{i=1}^{k} a_i^2 \leq v^{2-s/e}. \tag{4.3}$$

Then $p_0 = v^{-n/e}$ is a threshold probability for the appearance of $\mathcal{F}_b$.

Proof. Now, for each $\mathcal{F} \in \mathcal{F}_b$, let $A_{\mathcal{F}}$ be the event "$\mathcal{F}$ appears as a sub-hypergraph of $Q(n,p)$", and $X_{\mathcal{F}}$ be its indicator random variable. We have

$$\mathbb{P}(A_{\mathcal{F}}) = p^e,$$

as all e distinct sets of $\mathcal{F}$ lie in $Q(n,p)$. Set $X = \sum_{\mathcal{F}\in\mathcal{F}_b} X_{\mathcal{F}}$, the number of members of $\mathcal{F}_b$ that appear in $Q(n,p)$. By linearity, the expectation of X is

$$\mathbb{E}X = \mathbb{E}\sum_{\mathcal{F}\in\mathcal{F}_b} X_{\mathcal{F}} = \sum_{\mathcal{F}\in\mathcal{F}_b} \mathbb{E}X_{\mathcal{F}} = \sum_{\mathcal{F}\in\mathcal{F}_b} \mathbb{P}(A_{\mathcal{F}}) = \sum_{\mathcal{F}\in\mathcal{F}_b} p^e = \Theta(v^n p^e),$$

after using Equation 4.2.

When $p = o(p_0) = o(v^{-n/e})$, this quantity is approaching 0 as $n \to \infty$. Hence, $Q(n,p)$ excludes the appearance of $\mathcal{F}_b$ almost surely.

Now, two events $A_{\mathcal{F}_1}$ and $A_{\mathcal{F}_2}$ are not independent (denoted by $\mathcal{F}_1 \sim \mathcal{F}_2$) if and only if $\mathcal{F}_1$ and $\mathcal{F}_2$ share some common sets. Let $\mathcal{F}_i = \{F_1^{(i)}, \ldots, F_t^{(i)}\}$ for $i = 1, 2$, and let $|\mathcal{F}_1 \cap \mathcal{F}_2| = s$. $\mathcal{F}_1 \cap \mathcal{F}_2$ gives rise to two sub-hypergraphs $\mathcal{G}_1$ and $\mathcal{G}_2$ of $\mathcal{H}_b$, as the positions of these s sets in $\mathcal{F}_1$ and $\mathcal{F}_2$ could be different. Let $A_1, \ldots, A_k$ and $B_1, \ldots, B_k$ be the atoms of the σ-algebras generated by $\mathcal{G}_1$ and $\mathcal{G}_2$ on $V(\mathcal{H}_b)$ respectively, where A_i and B_i correspond to the same cell of the partition determined by the s common sets on $[n]$. It follows from Equation 4.1 that

$$\bigcup_{S\in A_i} F_S^{(1)} = \bigcup_{S\in B_i} F_S^{(2)}, \forall i = 1, \ldots, k.$$

Let $|A_i| = a_i, |B_i| = b_i$. It can be seen that a pair of dependent members of $\mathcal{F}_b$, with intersection of size s, and the positions of the s common sets fixed, correspond to an ordered partition of

$[n]$ into $\sum_{i=1}^{k} a_i b_i$ non-empty parts. Indeed, each $x \in [n]$ must be placed in an $F_S^{(1)}, S \in A_i$ and an $F_T^{(2)}, T \in B_i$, for some i. By the Cauchy–Schwartz inequality,

$$\sum_{i=1}^{k} a_i b_i \leq \sqrt{\sum_{i=1}^{k} a_i^2} \sqrt{\sum_{i=1}^{k} b_i^2} \leq \max\left(\sum_{i=1}^{k} a_i^2, \sum_{i=1}^{k} b_i^2\right) \leq v^{2-s/e}.$$

Taking the choice of positions of the s common sets in $\mathcal{F}_1$ and $\mathcal{F}_2$ into consideration, the number of these dependent pairs is

$$O\left(\sum_{s=1}^{e-1} s! \binom{e}{s}^2 v^{2n-sn/e}\right) = O\left(\sum_{s=1}^{e-1} v^{2n-sn/e}\right).$$

We calculate the following

$$\sum_{\mathcal{F}_1 \sim \mathcal{F}_2} \mathbb{P}(A_{\mathcal{F}_1} \wedge A_{\mathcal{F}_2}) = \sum_{\mathcal{F}_1 \sim \mathcal{F}_2} p^{2e-s} = O\left(\sum_{s=1}^{e-1} v^{2n-sn/e} p^{2e-s}\right).$$

Hence, when $p = \omega(p_0) = \omega(v^{-n/e})$, $\mathbb{E}X = \Theta(v^n p^e) \to \infty$, and

$$\frac{\sum_{\mathcal{F}_1 \sim \mathcal{F}_2} \mathbb{P}(A_{\mathcal{F}_1} \wedge A_{\mathcal{F}_2})}{(\mathbb{E}X)^2} = O\left(\sum_{s=1}^{e-1} v^{-ns/e} p^{-s}\right) = o(1).$$

Therefore, in this case, the second moment method (Corollary 4.3.4 in (85)) shows that $X > 0$ almost always, i.e. $Q(n,p)$ contains a member from $\mathcal{F}_b$ asymptotically almost surely. $\square$

4.3.2 Restricted second moment

We need some lemmas regarding the order of magnitude of multinomial coefficients in order to prove our next theorem.

Lemma 4.3.2. *Let α_i and $\lambda_i, i = 1, \ldots, k$ be positive numbers with $\sum_{i=1}^k \alpha_i = 1$. Then as $n \to \infty$, the set*

$$S = \{(n_i)_{i\in[k]} : \forall i, |n_i - \alpha_i n| \le \lambda_i \sqrt{n} \wedge \sum_{i=1}^{k} n_i = n\}$$

has cardinality $\Theta(n^{(k-1)/2})$.

Proof. To determine an element $(n_i)_{i\in[k]}$ of S. Let's first choose n_i one by one for i up to $k-1$. Each n_i has $2\lambda_i\sqrt{n}$ choices, yielding a total of $O(n^{(k-1)/2})$ choices. Now for n_k, there is at most 1 choice, since all the n_i's sum up to n. So we have proved that $|S| = O(n^{(k-1)/2})$.

To see the lower bound, we choose $\epsilon = \min_i \lambda_i/(k-1)$, and let

$$S' = \{(n_i)_{i\in[k]} : \forall i \in [k-1], |n_i - \alpha_i n| \le \epsilon\sqrt{n} \wedge |n_k - \alpha_k n| \le \lambda_k\sqrt{n} \wedge \sum_{i=1}^{k} n_i = n\}.$$

Then $\epsilon\sqrt{n} \le \lambda_i\sqrt{n}, \forall i \in [k-1]$, so $S' \subset S$. We count the elements of S'. For the first $k-1$ n_i's, the total number of choices is $(2\epsilon)^{k-1} n^{(k-1)/2} = \Omega(n^{(k-1)/2})$. All these choices are valid, since the deviation of n_i from $\alpha_i n$ is $O(\sqrt{n}) = o(n)$ so that the sum $\sum_{i=1}^{k-1} n_i \sim \sum_{i=1}^{k-1} \alpha_i n < n$ as $n \to \infty$. Moreover, for each such choice of n_i's we have

$$\left|\sum_{i=1}^{k-1} n_i - \sum_{i=1}^{k-1} \alpha_i n\right| \le (k-1)\epsilon\sqrt{n} = \min_i \lambda_i\sqrt{n} \le \lambda_k\sqrt{n}.$$

Thus there is always a choice for n_k, with $|n_k - \alpha_k n| \le \lambda_k \sqrt{n}$ and $\sum_{i=1}^k n_i = n$. This proves that $|S| \ge |S'| = \Omega(n^{(k-1)/2})$ which completes the proof. □

Lemma 4.3.3. *Let $\alpha_i, i \in [k]$ be positive reals with $\sum_{i=1}^k \alpha_i = 1$. If when $n \to \infty$, $n_i \sim \alpha_i n$ for all $i \in [k]$, then the multinomial coefficient*

$$\binom{n}{n_1, \ldots, n_k} = \Theta\left(\frac{1}{n^{(k-1)/2} \prod_{i=1}^k \alpha_i^{\alpha_i n}}\right).$$

Proof. By Stirling's formula,

$$\begin{aligned}\binom{n}{n_1, \ldots, n_k} &= \frac{n!}{\prod_{i=1}^k n_i!} = \Theta\left(\frac{\sqrt{n}(n/e)^n}{\prod_{i=1}^k \sqrt{n_i}(n_i/e)^{n_i}}\right) = \Theta\left(\frac{\sqrt{n}(n/e)^n}{(\prod_{i=1}^k \sqrt{n})(\prod_{i=1}^k (n_i/e)^{n_i})}\right)\\ &= \Theta\left(\frac{n^n/e^n}{n^{(k-1)/2}(\prod_{i=1}^k n_i^{n_i})/e^{\sum_{i=1}^k n_i}}\right) = \Theta\left(\frac{n^{\sum_{i=1}^k n_i}/e^n}{n^{(k-1)/2}(\prod_{i=1}^k n_i^{n_i})/e^n}\right)\\ &= \Theta\left(\frac{1}{n^{(k-1)/2}\prod_{i=1}^k (n_i/n)^{n_i}}\right) = \Theta\left(\frac{1}{n^{(k-1)/2}\prod_{i=1}^k \alpha_i^{\alpha_i n}}\right).\end{aligned}$$

□

Lemma 4.3.4. *Let α_i and $\lambda_i, i = 1, \ldots, k$ be positive reals with $\sum_{i=1}^k \alpha_i = 1$. Then, as $n \to \infty$,*

$$\sum_{\substack{|n_i - \alpha_i n| \le \lambda_i \sqrt{n}, \forall i \\ \sum_{i=1}^k n_i = n}} \binom{n}{n_1, \ldots, n_k} = \Theta\left(\frac{1}{\prod_{i=1}^k \alpha_i^{\alpha_i n}}\right).$$

Proof. For all $i \in [k]$, let n_i be such that $|n_i - \alpha_i n| \le \lambda_i\sqrt{n}$ and $\sum_{i=1}^k n_i = n$. Since the deviation of each n_i from $\alpha_i n$ is $O(\sqrt{n}) = o(n)$, when n tends to infinity, $n_i = (1+o(1))\alpha_i n$. So by Lemma 4.3.3,

$$\binom{n}{n_1,\ldots,n_k} = \Theta\left(\frac{1}{n^{(k-1)/2}\prod_{i=1}^k \alpha_i^{\alpha_i n}}\right).$$

Now, by Lemma 4.3.2, the total number of k-tuples $(n_1,\ldots,n_k)$ is on the order of $\Theta(n^{(k-1)/2})$. Summing up these multinomial coefficients, the $n^{(k-1)/2}$ on the denominator will be canceled out. This completes the proof. □

Lemma 4.3.5. *Let a_i be positive integers with $\sum_{i=1}^k a_i = v$. Then, as $n \to \infty$,*

$$\sum_{S_1}\sum_{S_2}\sum_{S_3}\binom{n}{n_1,\ldots,n_k}\prod_{i=1}^k\binom{n_i}{p_{i,1},\ldots,p_{i,a_i}}\prod_{i=1}^k\binom{n_i}{q_{i,1},\ldots,q_{i,a_i}} = O\left(v^n\prod_{i=1}^k a_i^{a_i n/v}\right),$$

where

$$S_1 = \{(n_i)_{i\in[k]} : \forall i \in [k], |n_i - a_i n/v| \le a_i\sqrt{n} \wedge \sum_{i=1}^k n_i = n\},$$
$$S_2 = \{((p_{i,j})_{j\in[a_i]})_{i\in[k]} : \forall i,j, |p_{i,j} - n/v| \le \sqrt{n} \wedge \forall i, \sum_{j=1}^{a_i} p_{i,j} = n_i\},$$
$$S_3 = \{((q_{i,j})_{j\in[a_i]})_{i\in[k]} : \forall i,j, |q_{i,j} - n/v| \le \sqrt{n} \wedge \forall i, \sum_{j=1}^{a_i} q_{i,j} = n_i\}.$$

Proof. First, notice that as n tends to infinity, $n_i = (1+o(1))a_i n/v, \forall i$, $p_{i,j} = (1+o(1))n_i/a_i = (1+o(1))n/v$ and $q_{i,j} = (1+o(1))n_i/a_i = (1+o(1))n/v, \forall i, j$. Using Lemma 4.3.3 (with $\alpha_i = n_i/n$ or $p_{i,j}/n_i$ or $q_{i,j}/n_i$ respectively), we get

$$
\begin{aligned}
&\binom{n}{n_1,\ldots,n_k}\prod_{i=1}^{k}\binom{n_i}{p_{i,1},\ldots,p_{i,a_i}}\prod_{i=1}^{k}\binom{n_i}{q_{i,1},\ldots,q_{i,a_i}}\\
=&\Theta\left(\frac{1}{n^{(k-1)/2}\prod_{i=1}^{k}(n_i/n)^{n_i}}\prod_{i=1}^{k}\frac{1}{n_i^{(a_i-1)/2}\prod_{j=1}^{a_i}(p_{i,j}/n_i)^{p_{i,j}}}\prod_{i=1}^{k}\frac{1}{n_i^{(a_i-1)/2}\prod_{j=1}^{a_i}(q_{i,j}/n_i)^{q_{i,j}}}\right)\\
=&\Theta\left(\frac{1}{n^{(k-1)/2}\prod_{i=1}^{k}(a_i/v)^{a_i n/v}}\prod_{i=1}^{k}\frac{1}{n^{(a_i-1)/2}\prod_{j=1}^{a_i}(1/a_i)^{n/v}}\prod_{i=1}^{k}\frac{1}{n^{(a_i-1)/2}\prod_{j=1}^{a_i}(1/a_i)^{n/v}}\right)\\
=&\Theta\left(\frac{\prod_{i=1}^{k}v^{a_i n/v}}{n^{(k-1)/2}\prod_{i=1}^{k}a_i^{a_i n/v}}\prod_{i=1}^{k}\frac{a_i^{2a_i n/v}}{n^{a_i-1}}\right)=\Theta\left(\frac{v^n}{n^{(k-1)/2}}\prod_{i=1}^{k}\frac{a_i^{a_i n/v}}{n^{a_i-1}}\right)\\
=&\Theta\left(\frac{v^n\prod_{i=1}^{k}a_i^{a_i n/v}}{n^{(k-1)/2}n^{v-k}}\right)=\Theta\left(\frac{v^n\prod_{i=1}^{k}a_i^{a_i n/v}}{n^{v-(k+1)/2}}\right).
\end{aligned}
$$

It remains to calculate the number of terms in the summation. By the proof of Lemma 4.3.2, we get bounds on sizes of S_i's. For S_1, we get directly from the lemma

$$|S_1| = O\left(n^{(k-1)/2}\right).$$

The reasoning for S_2 and S_3 is the same, so we only consider S_2. For each $i \in [k]$, the Lemma 4.3.2 shows that there are $O(n^{(a_i-1)/2})$ choices for the tuple $(p_{i,j})_{j\in[a_i]}$. So,

$$|S_2| = O\left(\prod_{i=1}^{k} n^{(a_i-1)/2}\right) = O\left(n^{(v-k)/2}\right).$$

So the total number of terms in the summation is

$$O\left(n^{(k-1)/2+2(v-k)/2}\right) = O(n^{v-(k+1)/2}).$$

Summing up the multivariate coefficients, we get the desired upper bound. $\square$

Consider the family of *almost-uniform-$\mathcal{F}_b$-configurations*, a restricted subfamily $\mathcal{F}_b^{au} \subset \mathcal{F}_b$, where a collection $F_1, F_2, \ldots, F_t$ is a member of $\mathcal{F}_b^{au}$ if $|F_S - n/v| \leq \sqrt{n}$ for all $S \subset [t]$ with $b_S = 1$. By proving the appearance of almost-uniform-$\mathcal{F}_b$-configurations, we get the same threshold function for $\mathcal{F}_b$.

Theorem 4.3.6 (Restricted Second Moment). *Fix $t \geq 2$ and a 0-1 vector $b = (b_S)_{S \subset [t]}$. Suppose that for all $\mathcal{G} \subset \mathcal{H}_b$ with $|\mathcal{G}| = s$ and $1 \leq s \leq e$,*

$$\prod_{i=1}^{k} a_i^{a_i/v} \leq v^{1-s/e}. \tag{4.4}$$

Then $p_0 = v^{-n/e}$ is a threshold probability for the appearance of $\mathcal{F}_b$.

Proof. The first moment argument as in the proof of Theorem 4.3.1 shows that when $p = o(p_0) = o(v^{-n/e})$, $Q(n,p)$ contains no members of $\mathcal{F}_b$, asymptotically almost surely. We consider the second moment. Observe that the members of $\mathcal{F}_b^{au}$ are in a one-to-one correspondence with ordered partitions of $[n]$ into v "almost equal" parts. Thus, by Lemma 4.3.4,

$$|\mathcal{F}_b^{au}| = \sum_{\substack{|n_i - n/v| \leq \sqrt{n}, \forall i \\ \sum_{i=1}^{v} n_i = n}} \binom{n}{n_1, \ldots, n_v} = \Theta\left(\frac{1}{\prod_{i=1}^{v} (1/v)^{n/v}}\right) = \Theta\left(v^n\right).$$

We use the notation for events and random variables as in the proof of Theorem 4.3.1. If $Y = \sum_{\mathcal{F}\in\mathcal{F}_b^{au}} X_{\mathcal{F}}$, then

$$\mathbb{E}Y = \mathbb{E}\sum_{\mathcal{F}\in\mathcal{F}_b^{au}} X_{\mathcal{F}} = \sum_{\mathcal{F}\in\mathcal{F}_b^{au}} \mathbb{E}X_{\mathcal{F}} = \sum_{\mathcal{F}\in\mathcal{F}_b^{au}} \mathbb{P}(A_{\mathcal{F}}) = \sum_{\mathcal{F}\in\mathcal{F}_b^{au}} p^e = \Theta\left(v^n p^e\right).$$

Consider two members $\mathcal{F}_1, \mathcal{F}_2$ of $\mathcal{F}_b^{au}$ with s sets in common. Again, these s sets partition $[n]$ into k parts, and we define the A_i's and B_i's in the same way as in the proof of Theorem 4.3.1. We have

$$\bigcup_{S\in A_i} F_S^{(1)} = \bigcup_{S\in B_i} F_S^{(2)}, \forall i = 1,\ldots,k.$$

But here $\mathcal{F}_1, \mathcal{F}_2$ are members of $\mathcal{F}_b^{au}$, so $||F_S^{(i)}| - n/v| \le \sqrt{n}$, for $i = 1, 2$, and S such that $b_S = 1$. So $a_i(n/v+\epsilon\sqrt{n}) = b_i(n/v+\epsilon'\sqrt{n}), \forall i \in [k]$, where $\epsilon, \epsilon' \in [-1, 1]$. As $n \to \infty$, this forces $a_i \approx b_i, \forall i$. Putting $\hat{a}_i = (a_i+b_i)/2$ we see that $a_i(n/v+\epsilon\sqrt{n}) = b_i(n/v+\epsilon'\sqrt{n}) = \hat{a}_i(n/v+\epsilon''\sqrt{n})$ where $\epsilon'' \in [-1, 1]$. Now these pairs $\mathcal{F}_1, \mathcal{F}_2$ are in a correspondence with ordered partitions of $[n]$ first into k parts of sizes $\hat{a}_i(n/v + \epsilon_i\sqrt{n}), i = 1,\ldots,k$, then the i-th part into $\hat{a}_i$ "almost equal" parts in two ways for each i. Hence when the s common sets are fixed, Lemma 4.3.5 calculates exactly the number of these pairs. The total number of dependent pairs is

$$O\left(\sum_{s=1}^{e-1} s!\binom{e}{s}^2 v^n \prod_{i=1}^{k} a_i^{a_i n/v}\right) = O\left(\sum_{s=1}^{e-1} v^{2n-sn/e}\right)$$

When $p = \omega(p_0) = \omega(v^{-n/e})$, such that $\mathbb{E}Y = \Theta(v^n p^e) \to \infty$, we have

$$\sum_{\mathcal{F}_1 \sim \mathcal{F}_2} \mathbb{P}(A_{\mathcal{F}_1} \wedge A_{\mathcal{F}_2}) = \sum_{\mathcal{F}_1 \sim \mathcal{F}_2} p^{2e-s} = O\left(\sum_{s=1}^{e-1} v^{2n-sn/e} p^{2e-s}\right).$$

Thus,

$$\frac{\sum_{\mathcal{F}_1 \sim \mathcal{F}_2} \mathbb{P}(A_{\mathcal{F}_1} \wedge A_{\mathcal{F}_2})}{(\mathbb{E}Y)^2} = O\left(\sum_{s=1}^{e-1} v^{-sn/e} p^{-s}\right) = o(1).$$

Therefore, the second moment method (Corollary 4.3.4 in (85)) shows that $Q(n,p)$ contains a member from $\mathcal{F}_b^{au}$, hence from $\mathcal{F}_b$, asymptotically almost surely. □

Let $\mathcal{F}_b^u$ be a more restricted subfamily of $\mathcal{F}_b$, which consists of all *uniform-$\mathcal{F}_b$-configurations*, that is, a collection $F_1, F_2, \ldots, F_t$ is a member of $\mathcal{F}_b^u$ if for all $S \subset [t]$ with $b_S = 1$, F_S are of the same size, namely n/v. By showing the appearance of uniform-$\mathcal{F}_b$-configurations, we get a weak threshold function for $\mathcal{F}_b$.

Theorem 4.3.7. ***(Another Restricted Second Moment)*** *Fix $t \geq 2$ and a 0-1 vector $b = (b_S)_{S \subset [t]}$. Suppose that condition* (Equation 4.4) *holds, then $p_0 = v^{-n/e}$ is a weak threshold probability for the appearance of $\mathcal{F}_b$.*

Proof. The first moment argument as in the proof of Theorem 4.3.1 shows that when $p = o(p_0) = o(v^{-n/e})$, $Q(n,p)$ contains no members of $\mathcal{F}_b$ asymptotically almost surely. Now, we are going to show that under the assumption of the theorem, that if p exceeds the threshold by n^r for a fixed sufficiently large r, then a member of $\mathcal{F}_b^u$ appears asymptotically almost surely.

We may assume again that $t = e$, and a similar argument shows that the members of $\mathcal{F}_b^u$ are in one-to-one correspondence with ordered partitions of $[n]$ into v equal parts. Using Lemma 4.3.3,

$$|\mathcal{F}_b^u| = \binom{n}{n/v, \ldots, n/v} = \Theta\left(\frac{v^n}{n^{(v-1)/2}}\right).$$

We use the notations for events and random variables as in the proof of Theorem 4.3.1. If $Z = \sum_{\mathcal{F} \in \mathcal{F}_b^u} X_{\mathcal{F}}$, then

$$\mathbb{E}Z = \mathbb{E} \sum_{\mathcal{F} \in \mathcal{F}_b^u} X_{\mathcal{F}} = \sum_{\mathcal{F} \in \mathcal{F}_b^u} \mathbb{E}X_{\mathcal{F}} = \sum_{\mathcal{F} \in \mathcal{F}_b^u} \mathbb{P}(A_{\mathcal{F}}) = \sum_{\mathcal{F} \in \mathcal{F}_b^u} p^e = \Theta\left(\frac{v^n p^e}{n^{(v-1)/2}}\right).$$

Consider two members $\mathcal{F}_1, \mathcal{F}_2$ of $\mathcal{F}_b^u$ with s sets in common. Again, the s sets partition $[n]$ into k parts, and we define A_i's, B_i's same as in the previous proof. Again, we have

$$\bigcup_{S \in A_i} F_S^{(1)} = \bigcup_{S \in B_i} F_S^{(2)}, \forall i = 1, \ldots, k. \tag{4.5}$$

But here $\mathcal{F}_1, \mathcal{F}_2$ are members of $\mathcal{F}_b^u$, so all $F_S^{(i)}$'s are of the same size n/v, for $i = 1, 2$, and S such that $b_S = 1$. So (Equation 4.5) implies that $a_i = b_i, \forall i \in [k]$. Now these pairs $\mathcal{F}_1, \mathcal{F}_2$ are in a correspondence with ordered partitions of $[n]$ first into k parts of sizes $a_i n/v, i = 1, \dots, k$, then the i-th part into a_i equal parts in two ways for each i. So the number of these pairs is

$$
\begin{aligned}
&O\left(\sum_{s=1}^{e-1} s!\binom{e}{s}^2 \binom{n}{a_1 n/v, \dots, a_k n/v}\left(\prod_{i=1}^{k}\binom{\frac{a_i n}{v}}{n/v, \dots, n/v}\right)^2\right)\\
=&O\left(\sum_{s=1}^{e-1} \frac{\sqrt{n}(n/e)^n}{\prod_{i=1}^k \sqrt{n}(a_i n/ev)^{a_i n/v}}\left(\prod_{i=1}^{k} \frac{\sqrt{n}(a_i n/ev)^{a_i n/v}}{(\sqrt{n}(n/ev)^{n/v})^{a_i}}\right)^2\right)\\
=&O\left(\sum_{s=1}^{e-1} \frac{v^n \prod_{i=1}^k a_i^{a_i n/v}}{n^{v-(k+1)/2}}\right) = O\left(\sum_{s=1}^{e-1} \frac{v^{2n-sn/e}}{n^{v-(k+1)/2}}\right)\\
=&O\left(\sum_{s=1}^{e-1} \frac{v^{2n-sn/e}}{n^{(v-1)/2}}\right).
\end{aligned}
$$

Note that the k in the above equations depends on the choice of s common sets, but $k \leq v$ always.

So taking $r = v$, when $p = \omega(p_0 n^v) = \omega(n^v v^{-n/e})$, $\mathbb{E}Z = \Theta\left(\frac{v^n p^e}{n^{(v-1)/2}}\right) \to \infty$ and

$$
\sum_{\mathcal{F}_1 \sim \mathcal{F}_2} \mathbb{P}(A_{\mathcal{F}_1} \wedge A_{\mathcal{F}_2}) = \sum_{\mathcal{F}_1 \sim \mathcal{F}_2} p^{2e-s} = O\left(\sum_{s=1}^{e-1} \frac{v^{2n-sn/e} p^{2e-s}}{n^{(v-1)/2}}\right).
$$

Thus,

$$
\frac{\sum_{\mathcal{F}_1 \sim \mathcal{F}_2} \mathbb{P}(A_{\mathcal{F}_1} \wedge A_{\mathcal{F}_2})}{(\mathbb{E}Z)^2} = O\left(\sum_{s=1}^{e-1} n^{(v-1)/2} v^{-sn/e} p^{-s}\right) = o(1).
$$

Therefore, the second moment method (Corollary 4.3.4 in (85)) shows that $Q(n, p)$ contains a member from $\mathcal{F}_b^u$, hence from $\mathcal{F}_b$, asymptotically almost surely. $\square$

Remark 4.3.8. *In the theorem above, if we were thinking of threshold functions instead of weak threshold functions, it wouldn't work. The expectation* $\mathbb{E}Z = \Theta\left(\frac{v^n p^e}{n^{(v-1)/2}}\right)$ *suggests that a threshold function seems like* $p_0 = n^{(v-1)/2e} v^{-n/e}$. *However, when* $p = \omega(p_0)$,

$$\frac{\sum_{\mathcal{F}_1 \sim \mathcal{F}_2} \mathbb{P}(A_{\mathcal{F}_1} \wedge A_{\mathcal{F}_2})}{(\mathbb{E}Z)^2} = O\left(\sum_{s=1}^{e-1} n^{(v-1)/2} v^{-sn/e} p^{-s}\right)$$

$$= O\left(\sum_{s=1}^{e-1} n^{(v-1)/2} v^{-sn/e} (n^{(v-1)/2e} v^{-n/e})^{-s}\right) = O\left(\sum_{s=1}^{e-1} n^{(1-s/e)(v-1)/2}\right).$$

which needs not to be infinitesimal.

4.4 A paradox

> "Horse" denotes form; "white" denotes color. What denotes color does not denote form. Therefore it is said, a white horse is not a horse. (86)
>
> —Kung-sun Lung-Tzu, circa 300 B.C.

It is worth mentioning that, to one's surprise, the hypothesis of the basic second moment theorem (Equation 4.3) is stronger than that of the restricted second moment theorem (Equation 4.4), which gives rise to paradoxical consequences.

Lemma 4.4.1. *Let* $t, v \geq 2$, $1 \leq s \leq e$, $a_i \geq 1$ *for* $i \in [k]$ *and* $\sum_i a_i = v$. *Suppose*

$$\sum_{i=1}^{k} a_i^2 \leq v^{2-s/e}.$$

Then

$$\prod_{i=1}^{k} a_i^{a_i/v} \leq v^{1-s/e}.$$

Proof. We use the weighted AM-GM inequality, which states that given positive numbers $x_1, \ldots, x_k$, and positive weights $w_1, \ldots, w_k$, the weighted arithmetic mean and the weighted geometric mean have the relation,

$$\prod_{i=1}^{k} x_i^{w_i/w} \leq \sum_{i=1}^{k} \frac{w_i x_i}{w},$$

where $w = \sum_i w_i$. Now, let $x_i = w_i = a_i$ for all $i \in [k]$, notice that $\sum_i a_i = v$, we have

$$\prod_{i=1}^{k} a_i^{a_i/v} \leq \sum_{i=1}^{k} \frac{a_i^2}{v} \leq \frac{v^{2-s/e}}{v} = v^{1-s/e}.$$

□

This means, when condition (Equation 4.3) holds, condition (Equation 4.4) must also hold. Thus if the probability p exceeds the threshold for condition (Equation 4.3) asymptotically, we are not only guaranteed that there exists an $\mathcal{F}_b$-configuration, but also guaranteed that there exists an almost-uniform-$\mathcal{F}_b$-configuration in our set system $Q(n,p)$.

Conversely, and probably more surprisingly, it is possible that for some configurations condition (Equation 4.4) holds while condition (Equation 4.3) fails to hold. If this is the case, as the probability p exceeds the threshold, it is inconclusive whether there is an $\mathcal{F}_b$-configuration according to Theorem 4.3.1, but Theorem 4.3.6 states that an almost-uniform-$\mathcal{F}_b$-configuration exists. In fact, this seemingly unusual phenomenon suggests that Theorem 4.3.1 can be improved, and the truth might be that both thresholds are always equal. This is a conjecture that we cannot prove right now.

4.5 Examples

In the light of the main theorems, finding the thresholds for the appearance of a specific configuration has been reduced to checking conditions (Equation 4.3) and (Equation 4.4). In this section, we consider some specific configurations, such as matchings, sunflowers and sequences of nested sets etc., calculate whether those conditions are satisfied. Thus, we get conclusions on threshold probabilities for their appearance.

4.5.1 Matchings

A *matching of size* t is a collection of t disjoint subsets $F_1, \ldots, F_t$ of $[n]$. The corresponding vector $b = (b_S)_{S \subset [t]}$ has $b_S = 1$ only when $S = \emptyset$ or $S = \{i\}$ for each $i \in [t]$. The graph $\mathcal{H}_b$ hence has $t+1$ vertices and t edges, all singletons. Take any subgraph $\mathcal{G}$ of $\mathcal{H}_b$, with $|\mathcal{G}| = s$. It's easy to see that $\mathcal{G}$ gives rise to a partition of $V(\mathcal{H}_b)$ into $k = s+1$ atoms, resulting $a_i = 1$ for $1 \leq i \leq s$ and $a_{s+1} = t - s + 1$.

We check the condition (Equation 4.3). Since,

$$\sum_{i=1}^{k} a_i^2 = s + (t-s+1)^2, \qquad v^{2-s/t} = (t+1)^{2-s/t}.$$

Suppose we take $s = t/2$, as $t \to \infty$, $\sum_{i=1}^{k} a_i^2 = \Theta(t^2)$, but $v^{2-s/t} = \Theta(t^{3/2})$. So the condition will be violated for t large enough. Actually, the condition (Equation 4.3) only holds for $t = 2, 3$.

However, condition (Equation 4.4) holds for any $t \geq 2$.

Theorem 4.5.1. *Given $t \geq 2$, let b be the vector corresponding to matchings of size t. Then, the threshold probability for the appearance of $\mathcal{F}_b$ is $(t+1)^{-n/t}$.*

Proof. It suffices to check condition (Equation 4.4), that is for any $1 \leq s \leq t$

$$\prod_{i=1}^{k} a_i^{a_i/v} \leq v^{1-s/t}.$$

From the discussion above, this is to show

$$\begin{aligned} & (t-s+1)^{(t-s+1)/(t+1)} \leq (t+1)^{1-s/t} \\ \Longleftrightarrow \quad & \frac{t-s+1}{t+1} \ln(t-s+1) \leq \left(1-\frac{s}{t}\right) \ln(t+1) \\ \Longleftrightarrow \quad & \left(1-\frac{s}{t}\right) \ln(t+1) - \frac{t-s+1}{t+1} \ln(t-s+1) \geq 0. \end{aligned}$$

Let the left side of the above be $f(s,t)$, we see that for any given t, $f(0,t) = f(t,t) = 0$. Moreover,

$$\frac{\partial^2}{\partial s^2} f(s,t) = -\frac{1}{(t+1)(t-s+1)} < 0,$$

as $s \in [0,t]$, which shows the graph of $f(s,t)$ (as a function in s) is concave on the interval $[0,t]$. So $f(s,t) \geq 0$. And by Theorem 4.3.6, the proof is complete. □

4.5.2 Sunflowers

A *sunflower with t petals* is a collection of t subsets $F_1, \ldots, F_t$ of $[n]$, with

- $F_i \cap F_j = \bigcap_{i=1}^{t} F_i \equiv C, \quad \forall i \neq j,$

- $F_i \setminus C \neq \emptyset, \quad \forall i.$

The definition implies that a matching of size t is meanwhile a sunflower with t petals (and an empty core). Hence, the discussion for matchings in the previous subsection also tells the threshold probability for sunflowers, since they are the same function of n. Now, more specifically, we are interested in sunflowers with non-empty cores:

- $C \neq \emptyset$.

With these restrictions, the corresponding vector b has coordinate $b_S = 1$ only where $S = \emptyset$ or $S = \{i\}$ for each $i \in [t]$ or $S = [t]$. The graph $\mathcal{H}_b$ is a star of size t (bipartite graph $K_{1,t}$), union with a single vertex. Take any subgraph $\mathcal{G}$ of $\mathcal{H}_b$, with $|\mathcal{G}| = s$. $\mathcal{G}$ is a star of size s union with $t-s+1$ isolated vertices. Hence, if $s \geq 2$, the atoms in the resulting partition have sizes $a_i = 1$ for $1 \leq i \leq s+1$ and $a_{s+2} = t-s+1$.

We check condition (Equation 4.3).

$$\sum_{i=1}^{k} a_i^2 = s + 1 + (t-s+1)^2 \text{ versus } v^{2-s/t} = (t+2)^{2-s/t}.$$

Again, take $s = t/2$, as $t \to \infty$, $\sum_{i=1}^{k} a_i^2 = \Theta(t^2)$, but $v^{2-s/t} = \Theta(t^{3/2})$. So the condition fails to hold for t large enough. Actually, the condition (Equation 4.3) only holds for $t = 2, 3, 4, \ldots, 9$.

However, we can again show that condition (Equation 4.4) holds.

Theorem 4.5.2. *Given $t \geq 2$, let b be the vector corresponding to sunflowers with t petals and non-empty cores. The threshold probability for the appearance of $\mathcal{F}_b$ is $(t+2)^{-n/t}$.*

Proof. It suffices to check condition (Equation 4.4), that is for any $1 \le s \le t$

$$\prod_{i=1}^{k} a_i^{a_i/v} \le v^{1-s/t}.$$

From the discussion above,

$$\begin{aligned} & (t-s+1)^{(t-s+1)/(t+2)} \le (t+2)^{1-s/t} \\ \Longleftrightarrow \quad & \frac{t-s+1}{t+2} \ln(t-s+1) \le \left(1-\frac{s}{t}\right) \ln(t+2) \\ \Longleftrightarrow \quad & \left(1-\frac{s}{t}\right) \ln(t+2) - \frac{t-s+1}{t+2} \ln(t-s+1) \ge 0. \end{aligned}$$

Let the left side of the above be $f(s,t)$, we see that for any given $t \ge 2$,

$$f(0,t) = \ln(t+2) - \frac{t+1}{t+2} \ln(t+1) \ge 0$$

and $f(t,t) = 0$. Moreover,

$$\frac{\partial^2}{\partial s^2} f(s,t) = -\frac{1}{(t+2)(t-s+1)} < 0,$$

as $s \in [0,t]$, which shows the graph of $f(s,t)$ (as a function in s) is concave on the interval $[0,t]$. So $f(s,t) \ge 0$. And by Theorem 4.3.6, the proof is complete. $\square$

4.5.3 Sequences of nested sets

A *sequence of nested sets of length* t is a collection of t subsets $F_1, \ldots, F_t$ of $[n]$, with

- $F_{i-1} \subset F_i, \quad \forall i \in [t]$,
- $F_i \setminus F_{i-1} \neq \emptyset, \quad \forall i \in [t]$. (Conventionally, we let $F_0 = \emptyset$.)

The vector $b = (b_S)_{S \subset [t]}$ has coordinate $b_S = 1$ only at where $S = \emptyset$ or $S = \{i, i+1, \ldots, t\}$ for each $i \in [t]$. Thus, the hypergraph $\mathcal{H}_b$ is also a sequence of t nested sets on $t+1$ vertices. Take any subgraph $\mathcal{G}$ of $\mathcal{H}_b$, with $|\mathcal{G}| = s$, say $\mathcal{G} = \{H_{i_j} : j \in [s]\}$, where $i_1 < i_2 < \ldots < i_s$. Then $\mathcal{G}$ is a sequence of nested sets of length s, which partitions $V(\mathcal{H}_b)$ into $k = s+1$ parts of sizes $a_j = i_j - i_{j-1}, \forall j \in [s+1]$ (setting $i_0 = 0$ and $i_{s+1} = t+1$ for convenience).

The condition (Equation 4.3) fails for large t's again. Since,

$$\sum_{i=1}^{k} a_i^2 = \sum_{j=1}^{s+1} (i_j - i_{j-1})^2, \qquad v^{2-s/t} = (t+1)^{2-s/t}.$$

We take $s = t/2$ and $i_j = j, \forall j \in [s]$, as $t \to \infty$, $\sum_{i=1}^{k} a_i^2 = \Theta(t^2)$, but $v^{2-s/t} = \Theta(t^{3/2})$, which makes the inequality impossible to hold. Actually, the condition (Equation 4.3) only holds for $t = 2, 3$.

Again, we can prove the threshold via Theorem 4.3.6. But we need a technical lemma about convex functions first.

Lemma 4.5.3. *Let $f(x)$ be a differentiable non-decreasing strictly convex function defined on positive real numbers. Let k be a positive integer, a, b be fixed positive numbers with $b \geq ak$. If $x_i \geq a$ for all $i \in [k]$ and $\sum_{i=1}^{k} x_i = b$, then $\sum_{i=1}^{k} f(x_i)$ is maximized at $x_1 = \ldots = x_{k-1} = a$ and $x_k = b - (k-1)a$.*

Proof. This follows from the fact that a convex function defined on a bounded polyhedron, is maximized at a vertex. $\square$

Theorem 4.5.4. *Given $t \geq 2$, let b be the vector corresponding to sequences of nested sets of length t, the threshold probability for the appearance of $\mathcal{F}_b$ is $(t+1)^{-n/t}$.*

Proof. It suffices to check condition (Equation 4.4), that is for any $1 \leq s \leq t$

$$\prod_{i=1}^{k} a_i^{a_i/v} \leq v^{1-s/t}.$$

From the discussion above, this is to show

$$\prod_{j=1}^{s+1} (i_j - i_{j-1})^{(i_j - i_{j-1})/(t+1)} \leq (t+1)^{1-s/t}.$$

Note that $\sum_{j=1}^{s+1}(i_j - i_{j-1}) = i_{s+1} - i_0 = t+1$, fix s and t, we maximize the left hand side with respect to values of $i_j - i_{j-1}$'s. To do so, take the logarithm,

$$\ln\left(\prod_{j=1}^{s+1} (i_j - i_{j-1})^{(i_j - i_{j-1})/(t+1)}\right) = \sum_{j=1}^{s+1} \frac{i_j - i_{j-1}}{t+1} \ln(i_j - i_{j-1})$$
$$= \frac{1}{t+1} \sum_{j=1}^{s+1} (i_j - i_{j-1}) \ln(i_j - i_{j-1}).$$

In Lemma 4.5.3, let $f(x) = x \ln x, k = s+1, a = 1, b = t+1$, then all the conditions are satisfied. So we see that the maximum is achieved when s of $i_j - i_{j-1}$'s are equal to 1, and one of them is equal to $t-s+1$. The original goal of maximization is also achieved as its logarithm's is. This shows

$$\prod_{j=1}^{s+1}(i_j - i_{j-1})^{(i_j - i_{j-1})/(t+1)} \leq (t-s+1)^{(t-s+1)/(t+1)}.$$

Now it suffices to show that

$$(t-s+1)^{(t-s+1)/(t+1)} \leq (t+1)^{1-s/t}.$$

But this has already been done in the case of matchings.

□

4.5.4 Configurations with $t \leq 4$

We list computational results in the appendix.

CHAPTER 5

THE LARGEST INTERSECTING FAMILY IN THE RANDOM SET SYSTEM

5.1 Notations and terminologies

Throughout this chapter, we will write $\mathcal{F} := \{\{A, B\} : A, B \subset [n] \text{ and } A \cap B = \emptyset\}$. With the notation for Turán problems defined in Chapter 1, the main purpose of this chapter is to show that

$$\mathrm{ex}(Q(n,p), \mathcal{F}) = (1 + o(1))p2^{n-1},$$

for p asymptotically larger than the threshold probability $p_0 = 2^{-\sqrt{n}\log n}$. We will apply a result due to Gauy, Hàn and Oliveira (68) regarding the random Erdős–Ko–Rado theorem. In order to do so, we follow the notion of supersaturation presented in their paper.

Definition 5.1.1. *Given $\lambda \in (0, 1], \gamma \in (0, 1]$, and a graph G on N vertices, we say that G is* (λ, γ)-supersaturated *if for any subset $S \subset V(G)$ with $|S| \geq \lambda N$, we have*

$$e(S) \geq \gamma \left(\frac{|S|}{N}\right)^2 e(G).$$

*In addition, let $\lambda = \lambda(n) > 0$ and $\gamma = \gamma(n) > 0$. A sequence $\{G_n\}_{n\in\mathbb{N}}$ is $(\lambda(n), \gamma(n))$-*supersaturated *if G_n is $(\lambda(n), \gamma(n))$-supersaturated for each $n \in \mathbb{N}$.*

The graph whose supersaturation property is of interest to us is an analogy to the Kneser graph. Let $G = (V, E)$ be the graph on the vertex set $V = 2^{[n]}$, and a pair of sets $\{A, B\} \in E$ if $A \cap B = \emptyset$. In other words, $E(G) = \mathcal{F}$. It follows from some simple calculation that the number of edges $|E(G)| = (3^n + 1)/2$ and hence its average degree $D = D(n) = (3^n + 1)/2^n$.

5.2 Supersaturation result for intersecting families

Note that the maximum intersecting family in $2^{[n]}$ has size 2^{n-1}. We prove in this section that even when the size of a family $\mathcal{A}$ exceeds this bound by an ϵ factor, there are "many" disjoint pairs in $\mathcal{A}$.

Theorem 5.2.1. *For all $\epsilon > 0$, let $\mathcal{A} \subset 2^{[n]}$ with $|\mathcal{A}| > (1 + \epsilon)2^{n-1}$. Then there exists a $\delta = \delta(\epsilon)$, such that the disjoint pairs in $\mathcal{A}$:*

$$|\mathcal{F} \cap \binom{\mathcal{A}}{2}| \geq 2^{n + \delta\sqrt{n}\log n}.$$

Proof. Thanks to a result due to Frankl (87) and Ahlswede (88), the number of disjoint pairs is minimized by a family $\mathcal{A}$ such that if $F, G \in 2^{[n]}$ with $|F| < |G|$, then $F \in \mathcal{A}$ implies $G \in \mathcal{A}$. Since $|\mathcal{A}| > (1 + \epsilon)2^{n-1}$, we may assume that $\mathcal{A}$ contains the top $n/2 + c\sqrt{n}$ layers of sets in $2^{[n]}$, where $c = c(\epsilon)$ (cf. Lemma 2.2.4).

Now think of a set $S \in \mathcal{A}$, with $n/2 - c\sqrt{n} \leq |S| \leq n/2 - c\sqrt{n}/2$. By the construction of $\mathcal{A}$, any set $T \subset [n] \setminus S$ such that $|T| > |S|$ will be in $\mathcal{A}$. Hence the number of sets in $\mathcal{A}$ that is disjoint from S is at least

$$\binom{n/2 + c\sqrt{n}/2}{n/2 - c\sqrt{n}/2} \geq 2^{\gamma\sqrt{n}\log n},$$

where the constant γ depends on c, hence depends on ϵ. Since the total number of sets in $\mathcal{A}$ of size in $[n/2 - c\sqrt{n}, n/2 - c\sqrt{n}/2]$ is at least $\delta' 2^n$, where δ depends on c (hence depends only on ϵ). We obtain that the total number of disjoint pairs in $\mathcal{A}$ is at least

$$\delta' 2^n \cdot 2^{\gamma\sqrt{n}\log n} \geq 2^{n+\delta\sqrt{n}\log n}$$

$\square$

It immediately follows from the theorem above that

Corollary 5.2.2. *Let n be an integer, $G = (V, E)$ be the graph on the vertex set $V = 2^{[n]}$, and a pair of sets $\{A, B\} \in E$ if $A \cap B = \emptyset$. Then, G is (λ, γ)-supersaturated, where $\lambda = (1+\epsilon)/2$ and*

$$\gamma = \left(\frac{2}{1+\epsilon}\right)^2 \cdot \frac{2^{n+\delta\sqrt{n}\log n + 1}}{3^n + 1}.$$

5.3 The largest intersecting family in the random set system

We state the main technical result due to Gauy, Hàn and Oliveira (68), which is an application of the container method. Given a graph H, let $\alpha(H)$ denote the size of the largest independent set in H. For a finite set V, let V_p be a random subset of V obtained by selecting each element $v \in V$ independently with probability p.

Theorem 5.3.1. *Proposition 2.6 of (68) (iii) Let $\lambda = \lambda(n)$ and $\gamma = \gamma(n)$ be $(0,1)$-valued functions, and let $G = \{G_n\}_{n\in\mathbb{N}}$ be a family of graphs, where each G_n has $N = N(n)$ vertices (with $\lim_{n\to\infty} N(n) = \infty$) and average degree $D = D(n)$. For any constant $0 < \epsilon < 1$, there exist a constant $C = C(\epsilon) > 0$ such that for any probability sequence $p = p(n) \in (0,1]$, the following holds. For a random spanning subgraph $G_n[V_p]$, where $V = V(G_n)$, we have:*

If G is (λ,γ)-supersaturated and $p \geq C(\lambda\gamma D)^{-1}\ln^2(e/\lambda)$, then

$$\mathbb{P}(\alpha(G_n[V_p]) \geq (1+\epsilon)\lambda pN) \leq \exp(-\epsilon^2 p\lambda N/24).$$

Now, we can easily translate the result into our case. Let G_n be defined as in last section, i.e., the graph on $2^{[n]}$ with edge set the set of all disjoint pairs. Then $V_p = Q(n,p)$, and any independent set in $G_n[V_p]$ is an intersecting family in $Q(n,p)$. So, $\mathrm{ex}(n,\mathcal{F}) = \alpha(G_n[V_p])$.

We have already shown that G_n is (λ,γ)-supersaturated, where $\lambda = (1+\epsilon)/2$ and

$$\gamma = \left(\frac{2}{1+\epsilon}\right)^2 \cdot \frac{2^{n+\delta\sqrt{n}\log n+1}}{3^n+1}.$$

Therefore, if

$$p \geq C(\lambda\gamma D)^{-1}\ln^2(e/\lambda) = C'/2^{\delta\sqrt{n}\log n},$$

then,

$$\mathbb{P}(\alpha(G_n[V_p]) \geq (1+\epsilon)\lambda pN) \leq \exp(-\epsilon^2 p\lambda N/24)$$

i.e.

$$\mathbb{P}(\mathrm{ex}(n,\mathcal{F}) \geq (1+\epsilon)^2 p 2^{n-1}) \leq \exp(-\epsilon^2 C' 2^{-\delta\sqrt{n}\log n}(1+\epsilon)2^n/48).$$

Also notice that with high probability, the largest intersecting family in $Q(n,p)$ is at least $(1+o(1))p2^{n-1}$. Therefore we have complete the proof of

Theorem 5.3.2. *Let $\mathcal{F} = \{\{A,B\} : A, B \subset [n]$ and $A \cap B = \emptyset\}$. If $p = 2^{-o(\sqrt{n}\log n)}$, then as $n \to \infty$*

$$\mathrm{ex}(Q(n,p),\mathcal{F}) = (1+o(1))p2^{n-1}.$$

Finally, we prove the threshold probability $p_0 = 2^{-\sqrt{n}\log n}$ is sharp. We state the following lemma whose standard proof can be found in many treaties of probability theory.

Lemma 5.3.3 (Hoeffding's inequality)**.** *Let X be a random variable that has a binomial distribution $B(n,p)$. Then*

$$\mathbb{P}(X < (p-\epsilon)n) < \exp(-2\epsilon^2 n),$$

and

$$\mathbb{P}(X > (p+\epsilon)n) < \exp(-2\epsilon^2 n).$$

Theorem 5.3.4. *Let $\mathcal{F} = \{\{A,B\} : A, B \subset [n]$ and $A \cap B = \emptyset\}$. If $p = 2^{-\Omega(\sqrt{n}\log n)}$, then there exists a constant $\epsilon > 0$ such that*

$$\mathrm{ex}(Q(n,p),\mathcal{F}) > (1+\epsilon)p2^{n-1},$$

with high probability.

Proof. Let $Q = Q(n,p)$. We want to show that for $p \ll 2^{-\sqrt{n}\log n}$, with high probability, one can find a intersecting family that is much larger than $(1+\epsilon)p2^{n-1}$ in Q, which establish the lower bound.

Define $Q_1 = Q \cap \binom{[n]}{>n/2}$. A set $S \subset [n]$ is said to be *good* if

- $n/2 - c\sqrt{n} < |S| \leq n/2$,
- $1 \in S$, and
- $S \cap T \neq \emptyset$, for all $T \in Q_1$.

Let $Q_0 = \{S \subset [n] : n/2 - c\sqrt{n} < |S| \leq n/2, 1 \in S,\}, Q_2 = Q \cap Q_0$, and $Q_3 = \{S \in Q : S \text{ is good}\}$.

By definition, $Q_1 \cup Q_3$ is an intersecting family in $Q = Q(n,p)$. We will show a lower bound on its size. Let $S \in Q_0$, i.e. S is a set that satisfies the first two conditions of goodness. Then, S is good if and only if all the sets with size greater than $n/2$ that is disjoint with S do not appear in Q. Let a be the number of such sets, we have

$$\begin{aligned} a &\leq c\sqrt{n}\binom{n/2 + c\sqrt{n}}{n/2} = c\sqrt{n}\binom{n/2 + c\sqrt{n}}{c\sqrt{n}} \\ &\leq c\sqrt{n}\left(\frac{e(n/2 + c\sqrt{n})}{c\sqrt{n}}\right)^{c\sqrt{n}} \leq 2^{c'\sqrt{n}\log n}. \end{aligned}$$

By independence, we have

$$\begin{aligned}\mathbb{P}\left(S \text{ is good}|S \in Q_0\right) &\geq (1-p)^a \geq (1-p)^{2^{c'\sqrt{n}\log n}} \\ &\geq (1-\epsilon)\exp(-p \cdot 2^{c'\sqrt{n}\log n}) \geq \frac{9}{10},\end{aligned}$$

where the last inequality holds since $p \ll 2^{-\sqrt{n}\log n}$ and ϵ is small. Hence,

$$\mathbb{P}\left(S \text{ is not good}|S \in Q_0\right) \leq \frac{1}{10}.$$

The probability that a bad set $S \in Q_0$ appears in Q (i.e. $S \in Q_2$) is then at most $p/10$. By Lemma 2.2.4, the number of sets that satisfies $|Q_0| = (1-\delta)2^{n-2}$, where $\delta = \delta(c)$. We get the following estimate on the expected number of bad sets

$$\mathbb{E}\left(\# \text{ of bad } S \in Q|S \in Q_0\right) \leq \frac{p}{10}(1-\delta)2^{n-2},$$

By Markov's inequality,

$$\mathbb{P}\left(\# \text{ of bad } S \in Q > \frac{kp}{10}(1-\delta)2^{n-2} \middle| S \in Q_0\right) \leq 1/k.$$

Taking $k = 2/(1-\delta)$,

$$\mathbb{P}\left(\# \text{ of bad } S \in Q > \frac{p}{5}2^{n-1} \middle| S \in Q_0\right) \leq \frac{1-\delta}{2} < 1/2.$$

Therefore, we conclude that, with probability at least $1/2$, the number of bad sets $S \in Q$ with $S \in Q_0$ is at most $(p/5)2^{n-1}$.

Now, let $X = |Q_1 \cup Q_2|$ denote the number of sets S such that either $|S| > n/2$ or $n/2 - c\sqrt{n} < |S| \leq n/2$ and $1 \in S$ that appears in Q. It is easy to see that X satisfies a binomial distribution $B(2^{n-1} + (1-\delta)2^{n-2}, p)$. By Hoeffding's inequality (Lemma 5.3.3), we have

$$\mathbb{P}\left(X > (p-\epsilon)(2^{n-1} + (1-\delta)2^{n-2})\right) \geq 1 - \exp(-2\epsilon^2(2^{n-1} + (1-\delta)2^{n-2})).$$

Let n be large enough, this probability is at least 0.9. Then with probability at least 0.4, both $|Q_1 \cup Q_2| = X > (p-\epsilon)(2^{n-1} + (1-\delta)2^{n-2}) = (p-\epsilon)(1.5 - \delta/2)2^{n-1} \geq 1.3p2^{n-1}$ and the number of bad $S \in Q_2$ is at most $(p/5)2^{n-1}$. Thus we obtain a large intersecting family $Q_1 \cup Q_3$ such that $|Q_1 \cup Q_3| \geq (1.1)p2^{n-1}$ with probability at least 0.4. This completes the proof.

$\square$

CITED LITERATURE

1. Euler, L.: Solutio problematis ad geometriam situs pertinentis. Commentarii academiae scientiarum Petropolitanae, 8:128–140, 1741.

2. Hierholzer, C. and Wiener, C.: Ueber die möglichkeit, einen linienzug ohne wiederholung und ohne unterbrechung zu umfahren. Mathematische Annalen, 6(1):30–32, Mar 1873.

3. Turán, P.: On an extremal problem in graph theory. Matematikai és Fizikai Lapok, 48:436–452, 1941. in Hungarian.

4. Erdős, P. and Rado, R.: Intersection theorems for systems of sets. Journal of London Mathematical Society, Second Series, 35(1):85–90, 1960.

5. Erdős, P. and Szemerédi, E.: Combinatorial properties of systems of sets. Journal of Combinatorial Theory, Series A, 24(3):308–313, 1978.

6. Alon, N., Shpilka, A., and Umans, C.: On sunflowers and matrix multiplication. Computational Complexity, 22:219–243, 2013.

7. Ellenberg, J. and Gijswijt, D.: On large subsets of $\mathbb{F}_q^n$ with no three-term arithmetic progression. Annals of Mathematics, 185:339–343, 2017.

8. Croot, E., Lev, V., and Pach, P.: Progression-free sets in $\mathbb{Z}_4^n$ are exponentially small. Annals of Mathematics, 185:331–337, 2017.

9. Naslund, E. and Sawin, W.: Upper bounds for sunflower-free sets. Forum of Mathematics, Sigma, 5:E15, 2017.

10. Erdős, P., Ko, C., and Rado, R.: Intersection theorems for systems of finite sets. Quarterly Journal of Mathematics: Oxford Journals, 12(1):313–320, 1961.

11. Hilton, A.: An intersection theorem for a collection of families of subsets of a finite set. Journal of London Mathematical Society, 2:369–384, 1977.

12. Pyber, L.: A new generalization of the erdős–ko–rado theorem. Journal of Combinatorial Theory, Series A, 43:85–90, 1986.

13. Matsumoto, M. and Tokushige, N.: The exact bound in the erdős–ko–rado theorem for cross-intersecting families. Journal of Combinatorial Theory, Series A, 52:90–97, 1989.

14. Bey, C.: On cross-intersecting families of sets. Graphs and Combinatorics, 21:161–168, 2005.

15. Borg, P.: Intersecting and cross-intersecting families of labeled sets. The Electronic Journal of Combinatorics, 15:#N9, 2008.

16. Frankl, P. and Tokushige, N.: On r-cross intersecting families of sets. Combinatorics, Probability and Computing, 20:749–752, 2011.

17. Frankl, P., Lee, S., Siggers, M., and Tokushige, N.: An erdős–ko–rado theorem for cross t-intersecting families. Journal of Combinatorial Theory, Series A, 128:207–249, 2014.

18. Keevash, P., Saks, M., Sudakov, B., and Verstraëte, J.: Multicolor turán problems. Advances in Applied Mathematics, 33:238–262, 2004.

19. Bollobás, B., Keevash, P., and Sudakov, B.: Multicolored extremal problems. Journal of Combinatorial Theory, Series A, 107:295–312, 2004.

20. Keevash, P. and Sudakov, B.: Set systems with restricted cross-intersections and the minimum rank of inclusion matrices. SIAM Journal on Discrete Mathematics, 18(4):713–727, 2005.

21. Frankl, P. and Rödl, V.: Forbidden intersections. Transactions of the American Mathematical Society, 300(1):259–286, 1987.

22. Mubayi, D. and Rödl, V.: Specified intersections. Transactions of the American Mathematical Society, 366(1):491–504, 2014.

23. Erdős, P., Kleitman, D., and B.L. RothschildSalam, A.: Asymptotic enumeration of k_n-free graphs. In Colloquio internazionale sulle teorie combinatorie, Vol. II, pages 19–27. Accademia nazionale dei Lincei.

24. Lamken, E. and Rothschild, B.: The numbers of odd-cycle-free graphs. Finite and infinite sets, Colloquia Mathematica Societatis János Bolyai, I (Eger, 1981)(37):547–553, 1984.

25. Balogh, J. and Samotij, W.: The number of $K_{m,m}$-free graphs. Combinatorica, 31:131–150, 2011.

26. Balogh, J. and Samotij, W.: The number of $K_{s,t}$-free graphs. Journal of London Mathematical Society, 83:368–388, 2011.

27. Balogh, J., Bollobás, B., and Simonovits, M.: The fine structure of octahedron-free graphs. Journal of Combinatorial Theory, Series B, 101(2):67–84, 2011.

28. Balogh, J., Bollobás, B., and Simonovits, M.: On the number of graphs without forbidden subgraph. Journal of Combinatorial Theory, Series B, 91:1–24, 2004.

29. Balogh, J., Bollobás, B., and Simonovits, M.: The typical structure of graphs without given excluded subgraphs. Random Structures Algorithms, 34(3):305–318, 2009.

30. Erdős, P., Frankl, P., and Rödl, V.: The asymptotic number of graphs not containing a fixed subgraph and a problem for hypergraphs having no exponent. Graphs and Combinatorics, 2:113–121, 1986.

31. Hundack, C., Prömel, H., and Steger, A.: Extremal graph problems for graphs with a color-critical vertex. Combinatorics, Probability and Computing, 2(4):465–477, 1993.

32. Prömel, H. and Steger, A.: The asymptotic number of graphs not containing a fixed color-critical subgraph. Combinatorica, 12(4):463–473, 1992.

33. Balogh, J. and Mubayi, D.: Almost all triple systems with independent neighborhoods are semi-bipartite. Journal of Combinatorial Theory, Series A, 118(4):1494–1518, 2011.

34. Balogh, J. and Mubayi, D.: Almost all triangle-free triple systems are tripartite. Combinatorica, 32(2):143–169, 2012.

35. Brightwell, G. and Goodall, S.: The number of partial orders of fixed width. Order, 13(4):315–337, 1996.

36. Brightwell, G., Grable, D., and Prömel, H. J.: Forbidden induced partial orders. Discrete Mathematics, 201(1–3):53–80, 1999.

37. Kleitman, D. and Rothschild, B.: Asymptotic enumeration of partial orders on a finite set. Transactions of the American Mathematical Society, 205:205–220, 1975.

38. Osthus, D., Kühn, D., Townsend, T., and Zhao, Y.: On the structure of oriented graphs and digraphs with forbidden tournaments or cycles. Journal of Combinatorial Theory, Series B, 124:88–127, 2017.

39. Person, Y. and Schacht, M.: Almost all hypergraphs without Fano planes are bipartite. In Proceedings of the Twentieth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA '09, pages 217–226, Philadelphia, PA, USA, 2009. Society for Industrial and Applied Mathematics.

40. Robinson, R.: Counting labeled acyclic digraphs. In New Directions in the Theory of Graphs, ed. F. Harary, pages 239–273. Academic Press.

41. Stanley, R.: Acyclic orientations of graphs. Discrete Mathematics, 5:171–178, 1973.

42. Nagle, B., Rödl, V., and Schacht, M.: Extremal hypergraph problems and the regularity method. In Topics in Discrete Mathematics, eds. M. Klazar, J. Kratochvíl, M. Loebl, J. Matoušek, P. Valtr, and R. Thomas, pages 247–278, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg.

43. Nagle, B. and Rödl, V.: The asymptotic number of triple systems not containing a fixed one. Discrete Mathematics, 235(1–3):271–290, 2001. Combninatorics (Prague, 1998).

44. Alon, A., Balogh, J., Bollobás, B., and Morris, R.: The structure of almost all graphs in a hereditary property. Journal of Combinatorial Theory, Series B, 101:85–110, 2011.

45. Kolaitis, P., Prömel, H., and Rothschild, B.: k_{l+1}-free graphs: asymptotic structure and a 0-1 law. Transactions of the American Mathematical Society, 303(2):637–671, 10 1987.

46. Conlon, D. and Gowers, W.: Combinatorial theorems in sparse random sets. Annual of Mathematics, 184:367–454, 2016.

47. Balogh, J., Morris, R., and Samotij, W.: Independent sets in hypergraphs. Journal of American Mathematical Society, 28:669–709, 2015.

48. Saxton, D. and Thomason, A.: Hypergraph containers. Inventiones Mathematicae, 201:1–68, 2015.

49. Bondy, J. and Simonovits, M.: Cycles of even length in graphs. Journal of Combinatorial Theory, Series B, 16:97–105, 1974.

50. Kővári, T., Sós, V., and Turán, P.: On a problem of k. zarankiewicz. Colloquium Mathematicum, 3:50–57, 1954.

51. Kleitman, D. and Winston, K.: On the number of graphs without 4-cycles. Discrete Mathematics, 41:167–172, 1982.

52. Kleitman, D. and Wilson, D.: On the number of graphs which lack small cycles. 1996. manuscript.

53. Morris, R. and Saxton, D.: The number of c_{2l}-free graphs. Advances in Mathematics, 298:534–580, 2016.

54. Balogh, J., Das, S., Delcourt, M., Liu, H., and Sharifzadeh, M.: Intersecting families of discrete structures are typically trivial. Journal of Combinatorial Theory, Series A, 132:224–245, 2015.

55. Balogh, J. and Wagner, A.: On the number of union-free families. Israel Journal of Mathematics, 219(1):431–448, 2017.

56. Balogh, J., Bohman, T., and Mubayi, D.: Erdős–ko–rado in random hypergraphs. Combinatorics, Probability and Computing, 18:629–646, 2009.

57. Bollobás, B., Narayanan, B., and Raigorodskii, A.: On the stability of the erdős–ko–rado theorem. Journal of Combinatorial Theory, Series A, 137:64–78, 2016.

58. Devlin, P. and Kahn, J.: On “stability” in the erdős–ko–rado theorem. SIAM Journal on Discrete Mathematics, 30(2):1283–1289, 2016.

59. Das, S. and Tran, T.: Removal and stability for erdős–ko–rado. SIAM Journal on Discrete Mathematics, 30(2):1102–1114, 2016.

60. Kostochka, A., Mubayi, D., and Verstraëte, J.: Turán problems and shadows i: paths and cycles. Journal of Combinatorial Theory, Series A, 129:57–79, 2015.

61. Kostochka, A., Mubayi, D., and Verstraëte, J.: Turán problems and shadows ii: trees. Journal of Combinatorial Theory, Series B, 122:457–478, 2017.

62. Kostochka, A., Mubayi, D., and Verstraëte, J.: Turán problems and shadows iii: expansions of graphs. SIAM Journal on Discrete Mathematics, 29(2):868–876, 2015.

63. Füredi, Z. and Jiang, T.: Hypergraph turán numbers of linear cycles. Journal of Combinatorial Theory, Series A, 123(1):252–270, 2014.

64. Füredi, Z., Jiang, T., and Seiver, R.: Exact solution of the hypergraph turán problem for k-uniform linear paths. Combinatorica, 34:299–322, 2014.

65. Erdős, P. and Rényi, A.: On the evolution of random graphs. Magyar Tudományos Akadémia Matematikai Kutató Intézetének Kőzleményei, 5:17–61, 1960.

66. Rényi, A.: On random subsets of a finite set. Mathematica (Cluj), 3(26):355–362, 1961.

67. Kreuter, B.: Small sublattices in random subsets of boolean lattices. Random Structures & Algorithms, 13(3-4):383–407.

68. Gauy, M., Hán, H., and Oliveira, I.: Erdős–ko–rado for random hypergraphs: asymptotics and stability. Combinatorics, Probability and Computing, 26(3):406–422, 2017.

69. Hamm, A. and Kahn, J.: On erdős–ko–rado for random hypergraphs i. arXiv, (1412.5085v1). preprint.

70. Hamm, A. and Kahn, J.: On erdős–ko–rado for random hypergraphs ii. arXiv, (1406.5793v3). preprint.

71. Balogh, J., Treglown, A., and Wagner, A.: Applications of graph containers in the boolean lattice. Random Structures & Algorithms, 49(4):845–872.

72. Borg, P.: The maximum sum and the maximum product of sizes of cross-intersecting families. European Journal of Combinatorics, 35:117–130, 2014.

73. Balogh, J., Morris, R., and Samotij, W.: The number of hypergraphs without linear cycles. Journal of Combinatorial Theory, Series B, (In press), 2018.

74. Ferber, A., McKinley, G., and Samotij, W.: Supersaturated sparse graphs and hypergraphs. International Mathematics Research Notices, page rny030, 2018.

75. Chung, F., Erdős, P., and Spencer, J.: On the decomposition of graphs into complete bipartite subgraphs. Studies in Pure Mathematics, pages 95–101, 1983. To the Memory of Paul Turán.

76. Bublitz, S.: Decomposition of graphs and monotone formula size of homogeneous functions. Acta Informatica, 23:689–696, 1986.

77. Tuza, Z.: Covering of graphs by complete bipartite subgraphs; complexity of 0-1 matrices. Combinatorica, 4(1):111–116, 1984.

78. Mubayi, D. and Turan, G.: Finding bipartite subgraphs efficiently. Information Processing Letters, 110(5, 1):174–177, 2010.

79. Bollobás, B.: Cycles modulo k. Bulletin of the London Mathematical Society, 9:97–98, 1977.

80. Thomassen, C.: Graph decomposition with applications to subdivisions and path systems modulo k. Journal of Graph Theory, 7:261–271, 1983.

81. Diwan, A.: Cycles of even lengths modulo k. Journal of Graph Theory, 65(3):246–252, 2010.

82. Liu, C.-H. and Ma, J.: Cycle lengths and minimum degree of graphs. Journal of Combinatorial Theory, Series B, 128:66–95, 2018.

83. Sudakov, B. and Verstraëte, J.: The extremal function for cycles of length l mod k. The Electronic Journal of Combinatorics, 22:#P00, 2015.

84. Collier-Cartaino, C., Graber, N., and Jiang, T.: Linear turán numbers of linear cycles and cycle-complete ramsey numbers. Combinatorics, Probability and Computing, 27(3):358–386, 2018.

85. Alon, N. and Spencer, J. H.: The Probabilistic Method. Hoboken, NJ, John Wiley & Sons, 2008. 3rd Edition.

86. Mei, Y. P.: The kung-sun lung-tzu with a translation into english. Harvard Journal of Asiatic Studies, 16(3/4):404–437, 1953.

87. Frankl, P.: On the minimum number of disjoint pairs in a family of finite sets. Journal of Combinatorial Theory, Series A, 22(2):249 – 251, 1977.

88. Ahlswede, R.: Simple hypergraphs with maximal number of adjacent pairs of edges. Journal of Combinatorial Theory, Series B, 28(2):164 – 167, 1980.

APPENDICES

Appendix A

COMPUTATIONAL RESULTS OF CHAPTER 4

A.1 Configurations with 3 sets

The coordinates of each vector b stand for the values of b_S with

$$S = \emptyset, \{1\}, \{2\}, \{3\}, \{1,2\}, \{1,3\}, \{2,3\}, \{1,2,3\}$$

respectively. For example, the vector $(1,1,1,1,0,0,0,1)$ gives rise to a sunflower with 3 petals. Conditions 1 and 2 refer to Equation 4.3 and Equation 4.4 respectively.

b	Condition 1	Condition 2	b	Condition 1	Condition 2
1,0,0,0,0,1,1,0	True	True	1,0,0,0,0,1,1,1	True	True
1,0,0,0,1,0,1,0	True	True	1,0,0,0,1,0,1,1	True	True
1,0,0,0,1,1,0,0	True	True	1,0,0,0,1,1,0,1	True	True
1,0,0,0,1,1,1,0	True	True	1,0,0,0,1,1,1,1	True	True
1,0,0,1,0,0,1,0	False	False	1,0,0,1,0,0,1,1	True	True
1,0,0,1,0,1,0,0	False	False	1,0,0,1,0,1,0,1	True	True
1,0,0,1,0,1,1,0	True	True	1,0,0,1,0,1,1,1	False	False
1,0,0,1,1,0,1,0	True	True	1,0,0,1,1,0,1,1	False	False
1,0,0,1,1,1,0,0	True	True	1,0,0,1,1,1,0,1	False	False

Appendix A (Continued)

1,0,0,1,1,1,1,0	True	True	1,0,0,1,1,1,1,1	False	True
1,0,1,0,0,0,1,0	False	False	1,0,1,0,0,0,1,1	True	True
1,0,1,0,0,1,1,0	True	True	1,0,1,0,0,1,1,1	False	False
1,0,1,0,1,0,0,0	False	False	1,0,1,0,1,0,0,1	True	True
1,0,1,0,1,0,1,0	True	True	1,0,1,0,1,0,1,1	False	False
1,0,1,0,1,1,0,0	True	True	1,0,1,0,1,1,0,1	False	False
1,0,1,0,1,1,1,0	True	True	1,0,1,0,1,1,1,1	False	True
1,0,1,1,0,0,0,0	False	False	1,0,1,1,0,0,0,1	True	True
1,0,1,1,0,0,1,0	False	False	1,0,1,1,0,0,1,1	False	False
1,0,1,1,0,1,0,0	True	True	1,0,1,1,0,1,0,1	False	False
1,0,1,1,0,1,1,0	False	False	1,0,1,1,0,1,1,1	False	False
1,0,1,1,1,0,0,0	True	True	1,0,1,1,1,0,0,1	False	False
1,0,1,1,1,0,1,0	False	False	1,0,1,1,1,0,1,1	False	False
1,0,1,1,1,1,0,0	False	False	1,0,1,1,1,1,0,1	True	True
1,0,1,1,1,1,1,0	False	True	1,0,1,1,1,1,1,1	True	True
1,1,0,0,0,1,0,0	False	False	1,1,0,0,0,1,0,1	True	True
1,1,0,0,0,1,1,0	True	True	1,1,0,0,0,1,1,1	False	False
1,1,0,0,1,0,0,0	False	False	1,1,0,0,1,0,0,1	True	True
1,1,0,0,1,0,1,0	True	True	1,1,0,0,1,0,1,1	False	False
1,1,0,0,1,1,0,0	True	True	1,1,0,0,1,1,0,1	False	False

Appendix A (Continued)

1,1,0,0,1,1,1,0	True	True	1,1,0,0,1,1,1,1	False	True
1,1,0,1,0,0,0,0	False	False	1,1,0,1,0,0,0,1	True	True
1,1,0,1,0,0,1,0	True	True	1,1,0,1,0,0,1,1	False	False
1,1,0,1,0,1,0,0	False	False	1,1,0,1,0,1,0,1	False	False
1,1,0,1,0,1,1,0	False	False	1,1,0,1,0,1,1,1	False	False
1,1,0,1,1,0,0,0	True	True	1,1,0,1,1,0,0,1	False	False
1,1,0,1,1,0,1,0	False	False	1,1,0,1,1,0,1,1	True	True
1,1,0,1,1,1,0,0	False	False	1,1,0,1,1,1,0,1	False	False
1,1,0,1,1,1,1,0	False	True	1,1,0,1,1,1,1,1	True	True
1,1,1,0,0,0,0,0	False	False	1,1,1,0,0,0,0,1	True	True
1,1,1,0,0,0,1,0	True	True	1,1,1,0,0,0,1,1	False	False
1,1,1,0,0,1,0,0	True	True	1,1,1,0,0,1,0,1	False	False
1,1,1,0,0,1,1,0	False	False	1,1,1,0,0,1,1,1	True	True
1,1,1,0,1,0,0,0	False	False	1,1,1,0,1,0,0,1	False	False
1,1,1,0,1,0,1,0	False	False	1,1,1,0,1,0,1,1	False	False
1,1,1,0,1,1,0,0	False	False	1,1,1,0,1,1,0,1	False	False
1,1,1,0,1,1,1,0	False	True	1,1,1,0,1,1,1,1	True	True
1,1,1,1,0,0,0,0	True	True	1,1,1,1,0,0,0,1	True	True
1,1,1,1,0,0,1,0	False	False	1,1,1,1,0,0,1,1	False	True
1,1,1,1,0,1,0,0	False	False	1,1,1,1,0,1,0,1	False	True

Appendix A (Continued)

1,1,1,1,0,1,1,0	False	False	1,1,1,1,0,1,1,1	True	True
1,1,1,1,1,0,0,0	False	False	1,1,1,1,1,0,0,1	False	True
1,1,1,1,1,0,1,0	False	False	1,1,1,1,1,0,1,1	True	True
1,1,1,1,1,1,0,0	False	False	1,1,1,1,1,1,0,1	True	True
1,1,1,1,1,1,1,0	True	True	1,1,1,1,1,1,1,1	True	True

A.2 Configurations with 4 sets

We have run a computer program to check all 4-set configurations. Since there are more than 30000 outputs, it is too long to display here.

A.3 Matchings, sunflowers and nested Sets

Have been investigated in the example section.

Appendix B

COPYRIGHT STATEMENT

B.1 Combinatorics, Probability, and Computing

The material in Chapter 2 has been previously published as *Multicolour Sunflowers* in Combinatorics, Probability, and Computing, 1–14. doi:10.1017/S0963548318000160, authored by Mubayi, D. and Wang, L. (2018). The publisher, the Cambridge University Press, has granted the author the permission to reuse the full article in this thesis (License No. 4375100049807).

B.2 Combinatorica

The material in Chapter 3 is a joint work with Dhruv Mubayi and will expect to appear in Combinatorica under the title *The number of triple systems without even cycles.* Springer allows the reuse of the material in a thesis by the author.

VITA

Education

Ph.D. in Mathematics, University of Illinois at Chicago, Chicago, Illinois, 2018

M.S. in Applied Mathematics, Illinois Institute of Technology, Chicago, Illinois, 2013

B.S. in Mathematics, Nankai University, Tianjin, China, 2010

Publications

Multicolour sunflowers, *Combinatorics, Probability and Computing,* 1–14. doi:10.1017/S0963548318000160, 2018, with D. Mubayi

The number of triple systems without even cycles, *Combinatorica* (accepted), arXiv:1701.00269v2, 2017, with D. Mubayi