UCLA UCLA Electronic Theses and Dissertations

Title Various Problems in Extremal Combinatorics

Permalink https://escholarship.org/uc/item/282049q8

Author Huang, Hao

Publication Date 2012

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Various Problems in Extremal Combinatorics

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Mathematics

by

Hao Huang

© Copyright by Hao Huang 2012

Abstract of the Dissertation

Various Problems in Extremal Combinatorics

by

Hao Huang

Doctor of Philosophy in Mathematics University of California, Los Angeles, 2012 Professor Benjamin Sudakov, Chair

Extremal combinatorics is a central theme of discrete mathematics. It deals with the problems of finding the maximum or minimum possible cardinality of a collection of finite objects satisfying certain restrictions. These problems are often related to other areas including number theory, analysis, geometry, computer science and information theory. This branch of mathematics has developed spectacularly in the past several decades and many interesting open problems arose from it. In this dissertation, we discuss various problems in extremal combinatorics, as well as some related problems from other areas.

This dissertation is organized in the way that each chapter studies a topic from extremal combinatorics, and includes its own introduction and concluding remarks. In Chapter 1 we study the relation between the chromatic number of a graph and its biclique partition, give a counterexample to the Alon-Saks-Seymour conjecture, and discuss related problems in theoretical computer science. Chapter 2 focuses on a conjecture on minimizing the number of nonnegative k-sums. Our approach naturally leads to an old conjecture by Erdős on hypergraph matchings. In Chapter 3, we improve the range that this conjecture is known to be true. Chapter 4 studies the connection of the Erdős conjecture with determining the minimum d-degree condition which guarantees the existence of perfect matching in hypergraphs. In Chapter 5, we study some extremal problems for Eulerian digraphs and obtain several results about existence of short cycles, long cycles, and subgraph with large minimum degree. The last chapter includes a proof that certain graph cut properties are quasi-random.

The dissertation of Hao Huang is approved.

Igor Pak

Bruce Rothschild

Amit Sahai

Benjamin Sudakov, Committee Chair

University of California, Los Angeles

2012

To my family, and friends

TABLE OF CONTENTS

1	Chr	comatic number and biclique partition	1	
	1.1	Introduction	1	
	1.2	Main Result	4	
	1.3	Neighborly families of boxes and t -biclique covering number $\ldots \ldots \ldots$	8	
	1.4	The clique vs. independent set communication problem	12	
	1.5	Concluding remarks	15	
2	Nor	nnegative k -sums, fractional covers and probability of small deviations	17	
	2.1	Introduction	17	
	2.2	Nonnegative k -sums and hypergraph matchings $\ldots \ldots \ldots \ldots \ldots \ldots$	21	
	2.3	Fractional covers and small deviations	24	
	2.4	Hilton-Milner type results	29	
	2.5	Concluding remarks	33	
3	The	e size of hypergraph and its matching number	35	
	3.1	Introduction	35	
	3.2	Shifting	38	
	3.3	Main result	40	
	3.4	Concluding Remarks	44	
4	Perfect Matching in hypergraphs			
	4.1	Introduction	46	
	4.2	Fractional matchings and probability of small deviations	52	
	4.3	Thresholds for perfect fractional matchings	56	

	4.4	Constructing integer matchings from fractional ones	57		
	4.5	An application in distributed storage allocation	64		
	4.6	Concluding Remarks	66		
5	\mathbf{Ext}	remal problems in Eulerian digraphs	68		
	5.1	Introduction	68		
	5.2	Feedback arc sets	71		
	5.3	Short cycles, long cycles, and Eulerian subgraphs with high minimum degree	79		
	5.4	Concluding remarks	83		
6	Qua	asi-randomness of graph balanced cut properties	86		
	6.1	Introduction	86		
	6.2	Preliminaries	89		
		6.2.1 Extremal Graph Theory	90		
		6.2.2 Concentration	92		
		6.2.3 Quasi-randomness of hypergraph cut properties	93		
	6.3	Base case - Triangle Balanced Cut	94		
	6.4	General Cliques	106		
	6.5	Concluding Remarks	112		
References					

Acknowledgments

First and foremost, I am deeply indebted to my Ph.D. advisor Professor Benjamin Sudakov, for his great patience and tireless guidance, and for his careful balancing between not pushing enough and pushing too much. It has always been a pleasant experience talking to him in all these years. His enthusiasm about mathematics and helpful advices are invaluable. My thanks also go to the rest of my committee members: Igor Pak, Bruce Rothschild, and Amit Sahai, for their tremendous support and guidance on finishing this dissertation.

For the one thousand and eight hundred days and nights of this long journey of Ph.D., I cannot imagine a better company than that of Yao Yao, who always tolerate my complaints and whining, give me solid support in various aspects of life, and always be the first audience of my buggy proofs.

I also want to give my sincere thanks to all my other collaborators and coauthors: Noga Alon, Shagnik Das, Peter Frankl, Choongbum Lee, Po-shen Loh, Jie Ma, Humberto Naves, Asaf Shapira, Vojtech Rödl, Andrzej Ruciński, and Raphy Yuster, for their inputs and contributions to our joint results. Without their help, it would be impossible for me to finish this dissertation. I also need to thank numerous other people who gave me many thoughtful career advices, and apologize that I cannot list all their names in a single page.

And last but not least my parents, Zhenqiu Huang and Xinhua Cheng, who introduced an ignorant kid into this magnificent world, watch him growing up without expecting any returns, and always unconditionally support his decisions and tolerate his stubbornness.

Vita

2007 B.S. (Mathematics), Peking University, China.

2007–2011 Teaching Assistant, Department of Mathematics, UCLA.

2011–2012 Dissertation Year Fellowship, UCLA.

PUBLICATIONS

Bandwidth theorem for random graphs (with C. Lee and B. Sudakov), Journal of Combinatorial Theory, Series B, 102 (2012), 14–37.

Quasi-randomness of graph balanced cut properties (with C. Lee), Random Structures & Algorithms, DOI: 10.1002/rsa.20384.

Nonnegative k-sums, fractional covers, and probability of small deviations (with N. Alon and B. Sudakov), Journal of Combinatorial Theory, Series B, 102 (2012), 784-796.

Large matchings in uniform hypergraphs and the conjectures of Erdős and Samuels (with N. Alon, P. Frankl, V. Rödl, A. Ruciński and B. Sudakov), *Journal of Combinatorial Theory*, *Series A*, 119 (2012), 1200-1215.

The size of a hypergraph and its matching number (with P. Loh and B. Sudakov), *Combinatorics, Probability and Computing*, 21 (2012), 442-450.

CHAPTER 1

Chromatic number and biclique partition

1.1 Introduction

Tools from linear algebra have many striking applications in the study of combinatorial problems. One of the earliest such examples is the theorem of Graham and Pollak [41]. Motivated by a communication problem that arose in connection with data transmission, they proved that the edge set of a complete graph K_k cannot be partitioned into disjoint union of less than k - 1 complete bipartite graphs. Their original proof used Sylvester's law of inertia. Over the years, this elegant result attracted a lot of attention and by now it has several different algebraic proofs, see [8, 77, 100, 103]. On the other hand, no purely combinatorial proof of this statement is known.

A natural generalization of Graham-Pollak theorem is to ask whether the same estimate holds also for all graphs with chromatic number k. This problem was raised twenty years ago by Alon, Saks, and Seymour, who made the following conjecture (see, e.g., the survey of J. Kahn [51]).

Conjecture 1.1.1. If the edges of a graph G can be partitioned into k edge disjoint complete bipartite graphs, then the chromatic number of G is at most k + 1.

This question is also related to another long-standing open problem by Erdős, Faber, and Lovász. They conjectured that the edge disjoint union of k complete graphs of order k is k-chromatic. Indeed, by replacing cliques in this problem by complete bipartite graphs we obtain the Alon-Saks-Seymour conjecture. The question of Erdős, Faber, and Lovász is still open. On the other hand, Kahn [50] proved the asymptotic version of their conjecture, showing that the chromatic number of edge disjoint union of k complete graphs of order k has chromatic number at most (1 + o(1))k.

Let $\mathbf{bp}(G)$ be the minimum number of bicliques (i.e., complete bipartite graphs) needed to partition the edges of graph G, and let $\chi(G)$ be the chromatic number of G. The Alon-Saks-Seymour Conjecture can be restated as $\mathbf{bp}(G) \geq \chi(G) - 1$. Until recently, there was not much known about this conjecture. Using the folklore result that the chromatic number of a union of graphs is at most the product of their chromatic numbers, one can easily get a lower bound $\mathbf{bp}(G) \geq \log_2 \chi(G)$. In [73], Mubayi and Vishwanathan improved the lower bound to $2\sqrt{2\log_2 \chi(G)}$. This estimate can be also deduced from the well known result of Yannakakis [104] in communication complexity. This connection to communication complexity was discovered by Alon and Haviv [4] (see Section 1.4 for details). Gao, McKay, Naserasr, and Stevens [38] introduced a reformulation of the Alon-Saks-Seymour conjecture and verified it for graphs with chromatic number $k \leq 9$. The main aim of this chapter is to obtain a superlinear gap between chromatic number and biclique partition number, which disproves the Alon-Saks-Seymour conjecture.

Theorem 1.1.2. There exist graphs G with arbitrarily large biclique partition number such that $\chi(G) \ge c \left(\mathbf{bp}(G)\right)^{6/5}$, for some fixed constant c > 0.

The study of (two-party) communication complexity, introduced by Yao [105], is an important topic in theoretical computer science which has many applications. In the basic model we have two players Alice and Bob who are trying to evaluate a boolean function $f: X \times Y \to \{0, 1\}$. Alice only knows x, Bob only knows y, and they want to communicate with each other according to some fixed protocol in order to compute f(x, y). The goal is to minimize the amount of communication during the protocol. The deterministic communication complexity D(f) is the number of bits that need to be exchanged for the worst inputs x, y by the best protocol for f. Let M be a matrix of f, i.e., $M_{x,y} = f(x, y)$ and let rk(M) be the rank of M. It's known that $D(f) \ge \log_2 rk(M)$. Lovász and Saks [66] conjectured that this bound is not very far from being tight. More precisely, their log-rank conjecture says that $D(f) \le (\log_2 rk(M))^{O(1)}$. This problem is directly related to the rank-coloring

conjecture of Van Nuffelen [102] and Fajtlowicz [31] in graph theory. This conjecture, which was disproved by Alon and Seymour [5], asked whether the chromatic number of a graph G is bounded by the rank of its adjacency matrix A_G . It is known that separation results between D(f) and $\log_2 rk(M)$ give corresponding separation between $\chi(G)$ and $rk(A_G)$. Several authors gave such separation results, e.g., [80, 81]. So far, the largest gap was obtained by Nisan and Wigderson [76] who constructed an infinite family of matrices such that $D(f) > (\log_2 rk(M))^{\log_2 3}$.

Similar to the rank-coloring problem, the Alon-Saks-Seymour conjecture is also closely related to a well known open problem in communication complexity. This communication problem, known as *clique versus independent set* (*CL-IS* for brevity), was introduced by Yannakakis [104] in 1988. In this problem, there is a publicly known graph G, Alice gets a clique C of G and Bob gets an independent set I of G. Their goal is to output $|C \cap I|$, which is clearly either 0 or 1. We will discuss the connection between this problem and the Alon-Saks-Seymour conjecture, and show that our counterexample yields the first nontrivial lower bound on the non-deterministic communication complexity of the *CL-IS* problem.

The rest of this chapter is organized as follows. In the next section we describe a counterexample to the Alon-Saks-Seymour Conjecture. In Section 1.3, we consider minimal coverings of a graph by bicliques, in which every edge of the graph is covered at least once and at most t times, for some parameter t. This more general notion is closely related to the question in combinatorial geometry about a neighborly family of boxes. We show that a natural variant of the Alon-Saks-Seymour conjecture for this more general parameter fails as well. In Section 1.4, we discuss connections with communication complexity and use our counterexample to obtain a new lower bound on the nondeterministic communication complexity of the clique vs. independent set problem. The final section contains some concluding remarks and open problems.

Notation. The *n*-dimensional cube Q_n is $\{0,1\}^n$ and two vertices x, y of Q_n are adjacent $x \sim y$ if and only if they differ in exactly one coordinate. A *k*-dimensional subcube of Q_n is a subset of $\{0,1\}^n$ which can be written as $\{x = (x_1, \dots, x_n) \in Q_n : x_i = a_i, \forall i \in T\}$, where

T is a set of n - k coordinates (called fixed coordinates), and each a_i is a fixed element in $\{0, 1\}$. In addition, we write 1^n and 0^n to represent the all-one and all-zero vectors in Q_n and use Q_n^- to indicate the set $Q_n \setminus \{1^n, 0^n\}$. Given two subsets $X \subset Q_k$ and $Y \subset Q_\ell$, we denote by $X \times Y$ the subset of the cube $Q_{k+\ell}$ which consists of all binary vectors (x, y) with $x \in X$ and $y \in Y$.

For a graph G = (V, E) with vertex set V and edge set E, we denote by $\chi(G)$, $\alpha(G)$, and **bp**(G) the chromatic number, independence number, and biclique partition number, respectively. The collection of all independent sets in G is denoted by $\mathcal{I}(G)$. Similarly, $\mathcal{C}(G)$ stands for the set of all cliques in G. The OR product of two graphs G and H is defined as a graph with vertex set equal to the Cartesian product $V(G) \times V(H)$, with adjacency $(g, h) \sim (g', h')$ iff $g \sim g'$ in G or $h \sim h'$ in H. The m-blowup of a graph G is obtained by replacing every vertex v of G with an independent set I_v of size m and by replacing every edge (u, v) of G with the complete bipartite graph whose parts are the independent sets I_u and I_v . We also use the notation $\mathcal{B}(U, W)$ to indicate the biclique with two parts U and W.

Throughout this chapter, we utilize the following standard notations to state asymptotic results. For two functions f(n) and g(n), write $f(n) = \Omega(g(n))$ if there exists a positive constant c such that $\liminf_{n\to\infty} f(n)/g(n) \ge c$, f(n) = o(g(n)) if $\limsup_{n\to\infty} f(n)/g(n) = 0$. Also, f(n) = O(g(n)) if there exists a strictly positive constant C > 0 such that $\limsup_{n\to\infty} f(n)/g(n) \le C$.

1.2 Main Result

In this section we describe a counterexample to the Alon-Saks-Seymour conjecture. Our construction is inspired by and somewhat similar to Razborov's counterexample to the rank-coloring conjecture [81]. Consider the following graph G = (V, E). Its vertex set is $V(G) = [n]^7 = \{(x_1, \dots, x_7) : x_i \in [n]\}$. For any two vertices $x = (x_1, \dots, x_7)$, $y = (y_1, \dots, y_7)$ in V(G), let ρ be the comparing function which records all coordinates

in which they differ. More precisely, $\rho(x, y) = (\rho_1(x, y), \cdots, \rho_7(x, y)) \in Q_7$, such that

$$\rho_i(x, y) = \begin{cases} 1 & \text{if } x_i \neq y_i \\ 0 & \text{if } x_i = y_i \end{cases}$$

Two vertices x and y are adjacent in G if and only if $\rho(x, y) \in S$, where S is the following subset of the cube Q_7

$$S = Q_7 \setminus \left[(1^4 \times Q_3^-) \cup \{0^4 \times 0^3\} \cup \{0^4 \times 1^3\} \right].$$

In the rest of this section we show that this graph G satisfies the assertion of Theorem 1.1.2.

Proposition 1.2.1. The independence number of G satisfies $\alpha(G) = O(n)$.

Proof. Let I be an independent set in G. For any set of indices $T = \{i_1, \ldots, i_t\} \subset \{1, 2, \cdots, 7\}$, let p_T be the natural projection of $[n]^7$ to $[n]^T$. More precisely, for every vector $x \in [n]^7$, p_T outputs the restriction of x to the coordinates in T, i.e., $p_T(x) = (x_{i_1}, \ldots, x_{i_t})$. For convenience, we will write p_{1234} instead of $p_{\{1,2,3,4\}}$, etc. If $|p_{1234}(I)| = 1$, it is easy to check from the definition of S, that any two vertices in I are different in each of the last three coordinates. As a result, $|I| = |p_{567}(I)| \leq n$. Now suppose that $|p_{1234}(I)| > 1$. Again from the definition of S, it follows that any two vertices $x, y \in G$ which agree on one of the first 4 coordinates and satisfy $p_{1234}(x) \neq p_{1234}(y)$ are adjacent in G. Hence, any two vectors in $p_{1234}(I)$ differ in all their coordinates and therefore $|p_{1234}(I)| \leq n$. If in addition, we also have for every element $x \in p_{1234}(I)$, $|p_{1234}^{-1}(x) \cap I| \leq 3$, then $|I| \leq 3|p_{1234}(I)| = O(n)$ and the proof is complete.

Otherwise, we may assume the existence of $\tilde{x} \in [n]^4$ and different vertices $\tilde{x}_1, \tilde{x}_2, \tilde{x}_3, \tilde{x}_4 \in I$ such that $p_{1234}(\tilde{x}_i) = \tilde{x}$. By the definition of S, it is easy to see that $p_{567}(\tilde{x}_i)$ differ in every coordinate. Since $|p_{1234}(I)| > 1$, there is a vertex $z \in I$ with $p_{1234}(z)$ different from \tilde{x} . Moreover, by the above discussion $p_{1234}(z)$ and \tilde{x} differ in every coordinate. As $1^7 \in S$, we also have that any two vertices of G which differ in all 7 coordinates are adjacent. This implies that $p_{567}(z)$ and $p_{567}(\tilde{x}_i)$ are equal in at least one coordinate. Since the number of coordinates of $p_{567}(z)$ is only 3 and there are 4 vertices $\tilde{x}_1, \tilde{x}_2, \tilde{x}_3, \tilde{x}_4$, we have that two of these vertices agree with $p_{567}(z)$ (and hence with each other) in the same coordinate. This contradicts the fact that $p_{567}(\tilde{x}_i)$ differ in all coordinates and completes the proof.

Corollary 1.2.2. The chromatic number of G is at least $\Omega(n^6)$.

Proof. Apply Proposition 1.2.1 together with the well-known fact that $\chi(G) \geq \frac{|V(G)|}{\alpha(G)}$. \Box

Proposition 1.2.3. The biclique partition number satisfies $bp(G) = O(n^5)$.

Before going into the details of the proof of this statement, we first need the following two lemmas.

Lemma 1.2.4. S can be partitioned into the disjoint union $S = \bigcup_{i=1}^{30} S_i$, where each S_i is a 2-dimensional subcube of Q_7 .

Proof. We start with the following simple observations.

(a) Q_3^- is a disjoint union of 1-dimensional subcubes.

(b) Q_3 can be decomposed into a disjoint union of 2-dimensional subcubes.

(c) For every $R_1 \subset Q_4$, the set $R_1 \times Q_3$ can be decomposed into a disjoint union of 2dimensional subcubes.

(d) For any $x_1 \sim x_2$ in Q_4 and $y_1 \sim y_2$ in Q_3 , the set $\{(x_1, y_1), (x_1, y_2), (x_2, y_1), (x_2, y_2)\}$ is a 2-dimensional subcube in Q_7 .

(e) For any $x_1 \sim x_2$ in Q_4 , $(x_1 \times Q_3^-) \cup (x_2 \times Q_3^-)$ can be decomposed into a disjoint union of 2-dimensional subcubes.

To verify (a), note that $Q_3^- = \{(0,0,1), (0,1,1)\} \cup \{(0,1,0), (1,1,0)\} \cup \{(1,0,0), (1,0,1)\}$. Claims (b) and (d) are obvious by the definition of a cube. Claim (c) is an immediate corollary of (b), and claim (e) follows easily from (a) and (d).

Next we can partition the set $S = Q_7 \setminus [(1^4 \times Q_3^-) \cup \{0^4 \times 0^3\} \cup \{0^4 \times 1^3\}]$ into the following 3 disjoint subsets S', S'', S''', and show that each of them is itself a disjoint union of 2-dimensional subcubes.

$$S' = \begin{cases} (0,0,0,0) \times Q_3^- \cup (0,0,0,1) \times Q_3^- \\ (0,0,1,1) \times Q_3^- \cup (1,0,1,1) \times Q_3^- \\ (0,1,0,1) \times Q_3^- \cup (0,1,1,1) \times Q_3^- \\ (1,1,0,1) \times Q_3^- \cup (1,0,0,1) \times Q_3^- \end{cases}$$

This set can be partitioned into a disjoint union of 2-dimensional subcubes, using claim (e).

$$S'' = \begin{cases} (1,1,1,1) \times 0^3 \cup (1,1,0,1) \times 0^3 \cup (1,0,1,1) \times 0^3 \cup (1,0,0,1) \times 0^3 \\ (1,1,1,1) \times 1^3 \cup (1,1,0,1) \times 1^3 \cup (1,0,1,1) \times 1^3 \cup (1,0,0,1) \times 1^3 \\ (0,1,1,1) \times 0^3 \cup (0,1,0,1) \times 0^3 \cup (0,0,1,1) \times 0^3 \cup (0,0,0,1) \times 0^3 \\ (0,1,1,1) \times 1^3 \cup (0,1,0,1) \times 1^3 \cup (0,0,1,1) \times 1^3 \cup (0,0,0,1) \times 1^3 \end{cases}$$

Note that every line in the definition of S'' describes a 2-dimensional subcube. This shows that S'' is a disjoint union of four 2-dimensional subcubes.

$$S''' = \begin{cases} (0,0,1,0) \times Q_3 \cup (0,1,0,0) \times Q_3 \cup (1,0,0,0) \times Q_3 \cup (0,1,1,0) \times Q_3 \\ (1,0,1,0) \times Q_3 \cup (1,1,0,0) \times Q_3 \cup (1,1,1,0) \times Q_3 \end{cases}$$

To decompose this set into a disjoint union of 2-dimensional subcubes, one can use claim (c).

Finally, it is easy to verify that indeed $S = S' \cup S'' \cup S'''$, and hence S can be partitioned into 2-dimensional subcubes.

Using the decomposition $S = \bigcup_{i=1}^{30} S_i$ from Lemma 1.2.4, we can define the following subgraphs $G_i \subset G$. The vertex set $V(G_i) = V(G)$ and two vertices $x, y \in G_i$ are adjacent if and only if $\rho(x, y) \in S_i$. From this definition, it is easy to see that G is the edge disjoint union of subgraphs G_i . Next we will show that every G_i has a small biclique partition number.

Lemma 1.2.5. $bp(G_i) \le n^5$.

Proof. Recall that the set S_i , which is used to define the edges of G_i , is a 2-dimensional subcube of Q_7 . Therefore there exists a set $T = \{t_1, \ldots, t_5\} \subset \{1, \cdots, 7\}$ of fixed coordinates

and $a_1, \ldots, a_5 \in \{0, 1\}$, such that $S_i = \{z = (z_1, \cdots, z_7) : z_{t_j} = a_j, \forall 1 \le j \le 5\}$. Also note that a_1, \ldots, a_5 can not all be simultaneously zero, since S does not contain 0^7 . Now we define the graph \widetilde{G}_i . Its vertex set $V(\widetilde{G}_i) = [n]^5$ and two vertices \widetilde{x} and \widetilde{y} are adjacent in \widetilde{G}_i if and only if $\rho(\widetilde{x}, \widetilde{y}) = (a_1, \ldots, a_5)$. It is rather straightforward to see that G_i is a n^2 -blowup of \widetilde{G}_i .

To complete the proof of this lemma we need two basic facts about the biclique partition number. The first one says that for any graph H, $\mathbf{bp}(H) \leq |V(H)| - 1$. Indeed, removing stars rooted at every vertex, one by one, we can partition every graph on h vertices into h-1 bicliques. The second one claims that if H is a blowup of \widetilde{H} , then $\mathbf{bp}(H) \leq \mathbf{bp}(\widetilde{H})$. To prove this, note that the blowup of a biclique is a biclique itself. Therefore the blowup of all the bicliques in a partition of \widetilde{H} becomes a biclique partition of H.

These two statements, together with the fact (mentioned above) that G_i is the blowup of \widetilde{G}_i , imply that $\mathbf{bp}(G_i) \leq \mathbf{bp}(\widetilde{G}_i) \leq |V(\widetilde{G}_i)| - 1 \leq n^5$. \Box

Proof of Proposition 1.2.3. Using that G is the edge disjoint union of G_i together with Lemma 1.2.5, we conclude that $\mathbf{bp}(G) = \mathbf{bp}(\bigcup_{i=1}^{30} G_i) \leq \sum_{i=1}^{30} \mathbf{bp}(G_i) = O(n^5).$

Propositions 1.2.2 and 1.2.3 show that the graph G, which we constructed, indeed satisfies the assertion of Theorem 1.1.2 and disproves the Alon-Saks-Seymour Conjecture.

1.3 Neighborly families of boxes and *t*-biclique covering number

The Alon-Saks-Seymour conjecture deals with the minimum number of bicliques needed to cover all the edges of a given graph G exactly once. It is also very natural to consider a more general problem in which we are allowed to cover the edges of graph at most t times. A *t*-biclique covering of a graph G is a collection of bicliques that cover every edge of G at least once and at most t times. The minimum size of such a covering is called the *t*-biclique covering number, and is denoted by $\mathbf{bp}_t(G)$. In particular, $\mathbf{bp}_1(G)$ is the usual biclique partition number $\mathbf{bp}(G)$.

In addition to being an interesting parameter to study in its own right, the t-biclique cov-

ering number is also closely related to a question in combinatorial geometry about neighborly families of boxes. A finite family C of d-dimensional convex polytopes is called t-neighborly if $d - t \leq \dim(C \cap C') \leq d - 1$ for every two distinct members C and C' of C. One particularly interesting case is when C consists of d-dimensional boxes with edges parallel to the coordinate axes. This type of box is called a *standard box*. Using the Graham-Pollak theorem, Zaks [106] proved that the maximum possible cardinality of a 1-neighborly family of standard boxes in \mathbb{R}^d is precisely d+1. His result was generalized by Alon [2], who proved that \mathbb{R}^d has a t-neighborly family of k standard boxes if and only if the complete graph K_k has a t-biclique covering of size d. This shows that the problem of determining the maximum possible cardinality of a t-neighborly family of standard boxes and the problem of computing the t-biclique covering number of a complete graphs are equivalent.

In his paper [2], Alon gave asymptotic estimates for $\mathbf{bp}_t(K_k)$, showing that

$$(1+o(1))(t!/2^t)^{1/t}k^{1/t} \le \mathbf{bp}_t(K_k) \le (1+o(1))t\,k^{1/t}$$

There is still a gap between these two bounds, and the problem of determining the right constant before $k^{1/t}$ is wide open even for the case t = 2. Using a different proof, we obtain here a slightly better lower bound of order roughly $(t!/2^{t-1})^{1/t}k^{1/t}$. For t = 2 it improves the above estimate by a factor of $\sqrt{2}$.

Proposition 1.3.1. If there exists a t-biclique covering of K_k of size d, then $k \leq 1 + \sum_{s=1}^{t} 2^{s-1} {d \choose s}.$

Proof. Suppose that the edges of K_k are covered by the bicliques $\{\mathcal{B}(U_j, W_j)\}_{j=1}^d$, such that every edge is covered at least once and at most *t*-times. For every nonempty subset of indices $S \subset [d]$ of size $|S| \leq t$ let $H_S = \bigcap_{j \in S} \mathcal{B}(U_j, W_j)$, and let A_S be the adjacency matrix of H_S . Let J be the $k \times k$ matrix of ones and let I be the $k \times k$ identity matrix. Then J - I is the adjacency matrix of K_k and it is easy to see, using the inclusion-exclusion principle, that

$$J - I = \sum_{S \subset [d], 0 < |S| \le t} (-1)^{|S| - 1} A_S$$

Also note that for |S| = s, the graph H_S is the disjoint union of at most 2^{s-1} smaller bicliques. Indeed, for every binary vector $z = (z_1, \ldots, z_{s-1})$, consider a complete bipartite graph with parts

$$X_z = \bigcap_{j,z_j=0} U_j \cap_{j,z_j=1} W_j \cap U_s \text{ and } Y_z = \bigcap_{j,z_j=0} W_j \cap_{j,z_j=1} U_j \cap W_s.$$

It is not difficult to check that these bicliques are disjoint and their union is H_S . Therefore, for every $S \subset [d], 0 < |S| = s \le t$ we can write $A_S = \sum_i B_{i,S}$, where $B_{i,S}$ is an adjacency matrix of a biclique and $1 \le i \le 2^{s-1}$. Thus we obtain that J - I can be written as a linear combination of at most $m = \sum_{s=1}^{t} 2^{s-1} {d \choose s}$ adjacency matrices of complete bipartite graphs.

Now to complete the proof we use the elegant trick of Peck [77] (we can use here other known proofs of the Graham-Pollak theorem as well). For the bipartite graph with adjacency matrix $B_{i,S}$, let $B'_{i,S}$ be the $k \times k$ matrix which contains only ones in positions whose row index lies in the first part of the bipartition and whose column index lies in the second part of the bipartition; the rest of the entries of $B'_{i,S}$ are zeros. Since the corresponding bipartite graph is complete, $B'_{i,S}$ has rank one. Furthermore, the matrix $B_{i,S} - 2B'_{i,S}$ is antisymmetric. As a result we can write J - I as a linear combination of at most m rank one matrices, plus some antisymmetric matrix T. Since an antisymmetric real matrix has only imaginary eigenvalues, I + T must have full rank k. But its rank can not exceed the rank of the linear combination of at most m rank one matrices plus J. As J has rank one as well, this implies that $k \leq m + 1 = 1 + \sum_{s=1}^{t} 2^{s-1} {d \choose s}$, which completes the proof.

As we already mentioned in the introduction, the motivation for the Alon-Saks-Seymour conjecture comes from the Graham-Pollak theorem which says that $\mathbf{bp}(K_k) \ge k-1$. Similarly, based on the lower bound of Alon that $\mathbf{bp}_t(K_k) \ge \Omega(k^{1/t})$, one can consider the following very natural generalization of this conjecture.

Question 1.3.2. Is it true that for every fixed integer t > 0, there exists a constant c = c(t)such that $\mathbf{bp}_t(G) \ge c(\chi(G))^{1/t}$ for all graphs G?

Recall that in Section 1.2 we constructed a graph G with $|V(G)| = n^7$ vertices such that $\alpha(G) = O(n)$ and $\mathbf{bp}(G) = O(n^5)$. Consider the OR product (defined in Section 1.1) of t

copies of G. We show that the graph G^t gives a negative answer to the above question for all positive integers t. This follows from the following sequence of claims.

Claim 1.3.3. $\alpha(G^t) \le \alpha(G)^t = O(n^t).$

Proof. We only need to prove $\alpha(G \times H) \leq \alpha(G)\alpha(H)$ for any two graphs G and H, since then the claim follows by induction on t. To prove this statement, consider a maximum independent set $I \in G \times H$. Let $I' = \{v \in G \mid (v, u) \in I \text{ for some } u \in H\}$ be the projection of I on V(G). By the definition of the OR product, this is an independent set in G and therefore has size at most $\alpha(G)$. Similarly, if I'' is the projection of I on V(H) then $|I''| \leq \alpha(H)$. To complete the proof, note that I is a subset of $I' \times I''$, and therefore its size cannot exceed $\alpha(G)\alpha(H)$.

Corollary 1.3.4. $\chi(G^t) = \Omega(n^{6t})$.

Proof. By Claim 1.3.3, $\chi(G^t) \ge \frac{|V(G^t)|}{\alpha(G^t)} \ge \frac{n^{7t}}{\alpha(G)^t} = \Omega(n^{6t}).$

Claim 1.3.5. $bp_t(G^t) \le t bp(G)$.

Proof. Consider graphs H_i , $1 \le i \le t$ with vertex set $V(H_i) = V(G^t)$, where two vertices (h_1, \dots, h_t) and (h'_1, \dots, h'_t) are adjacent in H_i if and only if $h_i \sim h'_i$ in G. Note that H_i is an n^{t-1} -blowup of G and therefore $\mathbf{bp}(H_i) = \mathbf{bp}(G)$. Also, it is easy to see that every edge in G^t is covered by some H_i . Since the number of graphs H_i is t, every edge of G^t is covered at most t times. Then the union of minimum biclique partitions of all H_i gives a t-biclique covering of G. Hence $\mathbf{bp}_t(G^t) \le \sum_{i=1}^t \mathbf{bp}(H_i) \le t \mathbf{bp}(G)$.

Claim 1.3.6. $bp_t(G^t) \leq c(\chi(G^t))^{\frac{5}{6t}}$ for some constant c = c(t).

Proof. By Claims 1.3.4 and 1.3.5, $\mathbf{bp}_t(G^t) \le t \mathbf{bp}(G) = O(tn^5) \le c(t) (\chi(G^t))^{\frac{5}{6t}}$.

This shows that the answer to Question 1.3.2 is negative for all natural t.

1.4 The clique vs. independent set communication problem

In Section 1.1, we defined the two-party communication model and discussed the concept of deterministic communication complexity. Here we need a few additional notions and definitions (see, e.g., [60] for more details). The non-deterministic communication complexity $N^1(f)$ of a function f is the smallest number of bits needed by an all powerful prover to convince Alice and Bob that f(x, y) = 1. It is known that $N^1(f) = \lceil \log_2 C^1(f) \rceil$, where $C^1(f)$ is the minimum number of monochromatic combinatorial rectangles needed to cover the 1inputs of the communication matrix M of f (recall that $M_{x,y} = f(x, y)$). With slight abuse of notation, we will later write $C^1(M)$ instead of $C^1(f)$. The numbers $N^0(f)$, $C^0(f)$, $C^0(M)$ are defined similarly, and the relation $N^0(f) = \lceil \log_2 C^0(f) \rceil$ holds as well.

In this section we consider the communication complexity of the clique versus independent set problem (CL-IS). In this problem, there is a publicly known graph Γ , Alice gets a clique C of Γ and Bob gets an independent set I of Γ . Their goal is to output $|C \cap I|$, which is clearly either 0 or 1. This problem was first introduced by Yannakakis [104], who also proposed the following algorithm to solve it. Given a graph Γ on m vertices, Alice sends to Bob the name of a vertex v in C whose degree in Γ is at most m/2. Note that in this case we can reduce the size of the graph by a factor of two by considering only the subgraph Γ' induced by the neighbors of v. In his turn, Bob sends Alice the name of a vertex u in his independent set $I \cap \Gamma'$ which has degree at least $|V(\Gamma')|/2$. In this case we can also reduce the size of the remaining problem by a factor of two. Finally if neither Alice nor Bob can send anything, it is easy to see that $C \cap I = \emptyset$. By repeating this procedure at most $\log_2 m$ rounds, one can show that the deterministic communication complexity satisfies $D(CL\text{-}IS_{\Gamma}) \leq O(\log_2^2 m)$. However, so far the best lower bound for this problem (see [59]) is only asymptotically $2\log_2 m$.

For the non-deterministic communication complexity of clique vs. independent set problem, it is easy to see that $N^1(CL-IS_{\Gamma})$ is always $\log m$. Indeed, for every vertex $v \in \Gamma$ consider the rectangle R_v formed by all cliques and all independent sets containing v. By definition, these m rectangles cover all 1-inputs of the communication matrix M of $CL-IS_{\Gamma}$. On the other hand, determining the correct order of magnitude of $N^0(CL-IS_{\Gamma})$ is wide open except for the trivial lower bound $\log_2 m$. This lower bound follows from the simple fact that taking all single vertices as cliques vs. the same vertices as independent sets shows that the $m \times m$ identity matrix is a submatrix of M. Next we discuss the connection between the Alon-Saks-Seymour conjecture and the CL-IS problem which was discovered by Alon and Haviv [4]. This connection together with our counterexample gives the first nontrivial lower bound for the nondeterministic communication complexity of the clique vs. independent set problem. It implies that there exists a graph Γ such that $N^0(CL-IS_{\Gamma}) \ge 6/5 \log_2 m - O(1)$.

Suppose we have a graph G = (V, E), V(G) = [n], $\mathbf{bp}(G) = m$, and a partition of E(G)into a disjoint union of bicliques $\{\mathcal{B}(U_i, W_i)\}_{i=1}^m$. Define the characteristic vector v_i of each biclique to be $v_i = (v_{i1}, \dots, v_{in}) \in \{0, 1, *\}^n$, so that

$$v_{ij} = \begin{cases} 0 & \text{if } j \in U_i \\ 1 & \text{if } j \in W_i \\ * & \text{otherwise} \end{cases}$$

Using the notations above, we create a new graph Γ with vertex set [m]. Two vertices iand i' are adjacent in Γ if there exists a $j \in [n]$ such that $v_{ij} = v_{i'j} = 1$. Two vertices i and i' are nonadjacent if there exists a $j' \in [n]$ such that $v_{ij'} = v_{i'j'} = 0$. In every other case, arbitrarily assign an edge or non-edge between i and i'. If there are two indices j, j' such that $v_{ij} = v_{i'j} = 1$ and $v_{ij'} = v_{i'j'} = 0$, then $j \in W_i \cap W_{i'}$ and $j' \in U_i \cap U_{i'}$. Therefore the edge (j', j) is covered by two bicliques, which is impossible since $\cup_{i=1}^m \mathcal{B}(U_i, W_i)$ is an edge partition of G. This shows that Γ is well defined.

Now consider the *CL-IS* problem on Γ . Define $C_j = \{q \in [m] : v_{qj} = 1\}$ and $I_j = \{q \in [m] : v_{qj} = 0\}$. By definition of Γ , it is easy to see that $\{C_j\}$ are cliques and $\{I_j\}$ are independent sets in this graph. Denote the matrix of *CL-IS*_{Γ} by *M*. Let *M'* be the submatrix of *M* corresponding to the rows determined by $\{C_j\}_{j=1}^n$ and columns determined by $\{I_j\}_{j=1}^n$. Obviously $N^0(M) \ge N^0(M') = \log_2 C^0(M')$. Assume that we have a covering of 0-entries of *M'* by monochromatic rectangles, and let R_1, \dots, R_t be the rectangles which cover the diagonal entries of *M'*. Note that $M'_{pp} = M'_{qq} = 0$ by definition. If M'_{pp} and M'_{qq}

are both covered by R_i , then $M'_{pq} = M'_{qp} = 0$ and thus $C_p \cap I_q$ and $C_q \cap I_p$ are both empty. This implies that (p,q) is not an edge in graph G, since otherwise there must exist an index i such that $v_{ip} = 0$, $v_{iq} = 1$ or $v_{ip} = 1$, $v_{iq} = 0$. Then either $i \in I_p \cap C_q$ or $i \in C_p \cap I_q$, which gives a contradiction. In particular, the family of rectangles $\{R_i\}_{i=1}^t$ corresponds to a covering of graph G by independent sets, and therefore $\chi(G) \leq t$. Thus we have that

$$N^{0}(M) \ge N^{0}(M') = \log_{2} C^{0}(M') \ge \log_{2} t \ge \log_{2} \chi(G).$$

This estimate, together with the existence of a graph G (from Section 1.2) which has $\mathbf{bp}(G) = O(\chi(G)^{5/6})$, proves the following theorem.

Theorem 1.4.1. There exists an infinite collection of graphs Γ , such that

$$N^{0}(CL-IS_{\Gamma}) \ge \frac{6}{5}\log_{2}|V(\Gamma)| - O(1).$$

In addition, the combination of the inequality $N^0(CL - IS_{\Gamma}) \ge \log_2 \chi(G)$ we just proved, and the result of Yannakakis that $D(CL - IS_{\Gamma}) \le O(\log_2^2 m)$, immediately gives a different derivation of the following result of Mubayi and Vishwanathan. It shows that if $\mathbf{bp}(G) = m$, then

$$\chi(G) \le 2^{N^0(CL - IS_{\Gamma})} \le 2^{D(CL - IS_{\Gamma})} \le 2^{O(\log_2^2 m)}.$$

From the above discussions, we know that any separation result between $\chi(G)$ and $\mathbf{bp}(G)$ gives corresponding separation between $N^0(CL-IS)$ and the trivial lower bound $\log_2 |V(\Gamma)|$. We do not yet know whether the converse is also true. However, a weaker converse does exist, as was observed by Alon and Haviv [4]. More precisely, the gap between $N^0(CL-IS_{\Gamma})$ and $\log_2 |V(\Gamma)|$ implies a gap between $\chi(H)$ and the 2-biclique covering number $\mathbf{bp}_2(H)$ for some graph H.

Let $\Gamma = (V, E)$ be a graph with vertices $V = \{v_1, \dots, v_m\}$ and consider the following graph H. The vertices of H are all the pairs (C, I) such that C is a clique and I is an independent set in Γ , and $C \cap I = \emptyset$. Two vertices (C, I) and (C', I') are adjacent if $C \cap I' \neq \emptyset$ or $C' \cap I \neq \emptyset$. For every vertex v_i in Γ , we define two subsets $U_i = \{(C, I) : v_i \in C\}$ and $W_i = \{(C, I) : v_i \in I\}$ of H. These subsets have the following properties.

- 1. U_i and W_i are disjoint.
- 2. (U_i, W_i) is a complete bipartite subgraph of H.
- 3. $G' = \bigcup_{i=1}^{m} \mathcal{B}(U_i, W_i)$ and each edge of H is covered at most two times.

Property (1) holds since $C \cap I = \emptyset$ for every vertex (C, I) of H. To verify (2), consider two vertices $(C, I) \in U_i$ and $(C', I') \in W_i$. Then $v_i \in C \cap I'$, which means $C \cap I' \neq \emptyset$ and thus (C, I) and (C', I') are adjacent in H. To prove (3), note that by definition, any edge $(C, I) \sim (C', I')$ in G' either satisfies $C \cap I' \neq \emptyset$ or $C' \cap I \neq \emptyset$ or both. If $C \cap I' \neq \emptyset$, then there is a unique i (since $|C \cap I'| \leq 1$) such that $v_i \in C$ and $v_i \in I'$, which means that this edge belongs to $\mathcal{B}(U_i, W_i)$. A similar conclusion holds in the case when $C' \cap I \neq \emptyset$. Thus every edge of H is covered by $\{\mathcal{B}(U_i, W_i)\}_{i=1}^m$ either once or twice. This shows that $\mathbf{bp}_2(H) \leq m = |V(\Gamma)|$.

Next we bound the chromatic number of H from below by a function of $N^0(CL \cdot IS_{\Gamma})$. Denote the matrix of $CL \cdot IS_{\Gamma}$ by M. By definition, an independent set $I' = \{(C_1, I_1), \dots, (C_l, I_l)\}$ in H corresponds to an all-zero submatrix of M, whose rows and columns are indexed by C_1, \dots, C_l and I_1, \dots, I_l respectively. Thus a proper coloring of H corresponds to a covering of the 0-entries of M by monochromatic rectangles. Therefore $\chi(H) \geq C^0(M) = C^0(CL \cdot IS_{\Gamma}) \geq 2^{N^0(CL \cdot IS_{\Gamma})}$, and hence we established the following claim.

Claim 1.4.2. For every graph Γ there exists a graph H such that

$$\mathbf{bp}_2(H) \le |V(\Gamma)|$$
 and $\chi(H) \ge 2^{N^0(CL-IS_{\Gamma})}$.

1.5 Concluding remarks

In this chapter we constructed a graph which has a polynomial gap between its chromatic number and its biclique partition number, thereby disproving the Alon-Saks-Seymour conjecture. A very interesting problem which remains widely open is to determine how large this gap can be. In communication complexity it is a long standing open problem to prove an $\Omega(\log^2 N)$ lower bound on the complexity of the clique vs. independent set problem for graphs on N vertices. Since, as we already explained in the previous section, this problem is closely related to the Alon-Saks-Seymour conjecture, it is plausible to believe that one can obtain a corresponding gap between chromatic and biclique partition numbers. We conjecture that there exists a graph G with biclique partition number k and chromatic number at least $2^{c \log^2 k}$, for some constant c > 0. The existence of such a graph will also resolve the complexity of the clique vs. independent set problem.

Another intriguing question which deserves further study is to determine the *t*-biclique covering numbers of complete graphs. This will also solve the problem of the maximum possible cardinalities of *t*-neighborly families of standard boxes in finite dimensional Euclidean spaces. Even the asymptotics of $\mathbf{bp}_t(K_k)$ are only known up to a multiplicative constant factor. In the first open case when t = 2, the current best bounds are $(1 + o(1))k^{1/2} \leq \mathbf{bp}_2(K_k) \leq (1 + o(1))2k^{1/2}$, and it would be interesting to close this gap.

Acknowledgment. The author would like to thank N. Alon for explaining to us his results with I. Haviv on the connection between the Alon-Saks-Seymour conjecture and the clique vs. independent set problem, P. Loh for very carefully reading the manuscript and many suggestions, and J. Fox, J. Greene, L. Lovász and B. Mohar for useful remarks.

CHAPTER 2

Nonnegative k-sums, fractional covers and probability of small deviations

2.1 Introduction

Let $\{x_1, \dots, x_n\}$ be a set of n real numbers whose sum is nonnegative. It is natural to ask the following question: how many subsets of nonnegative sum must it always have? The answer is quite straightforward, one can set $x_1 = n - 1$ and all the other $x_i = -1$, which gives 2^{n-1} subsets. This construction is also the smallest possible since for every subset A, either A or $[n] \setminus A$ or both must have a nonnegative sum. Another natural question is, what happens if we further restrict all the subsets to have a fixed size k? The same example yields $\binom{n-1}{k-1}$ nonnegative k-sums consisting of n-1 and (k-1) -1's. This construction is similar to the extremal example in the Erdős-Ko-Rado theorem [30] which states that for $n \ge 2k$, a family of subsets of size k in [n] with the property that every two subsets have a nonempty intersection has size at most $\binom{n-1}{k-1}$. However the relation between k-sum and k-intersecting family is somewhat subtle and there is no obvious way to translate one problem to the other.

Denote by A(n, k) the minimum possible number of nonnegative k-sums over all possible choices of n numbers x_1, \dots, x_n with $\sum_{i=1}^n x_i \ge 0$. For which values of n and k, is the construction $x_1 = n - 1, x_2 = \dots = x_n = -1$ best possible? In other words, when can we guarantee that $A(n, k) = {\binom{n-1}{k-1}}$? This question was first raised by Bier and Manickam [11, 12] in their study of the so-called first distribution invariant of the Johnson scheme. In 1987, Manickam and Miklós [68] proposed the following conjecture, which in the language of the Johnson scheme was also posed by Manickam and Singhi [69] in 1988.

Conjecture 2.1.1. For all $n \ge 4k$, we have $A(n,k) = \binom{n-1}{k-1}$.

In the Erdős-Ko-Rado theorem, if n < 2k, all the k-subsets form an intersecting family of size $\binom{n}{k} > \binom{n-1}{k-1}$. But for n > 2k, the star structure, which always takes one fixed element and k-1 other arbitrarily chosen elements, will do better than the set of all k-subsets of the first 2k - 1 elements. For a similar reason we have the extra condition $n \ge 4k$ in the Manickam-Mikós-Singhi conjecture. $\binom{n-1}{k-1}$ is not the best construction when n is very small compared to k. For example, take n = 3k+1 numbers, 3 of which are equal to -(3k-2) and the other 3k-2 numbers are 3. It is easy to see that the sum is zero. On the other hand, the nonnegative k-sums are those subsets consisting only of 3's, which gives $\binom{3k-2}{k}$ nonnegative k-sums. It is not difficult to verify that when k > 2, $\binom{3k-2}{k} < \binom{(3k+1)-1}{k-1}$. However this kind of construction does not exist for larger n.

The Manickam-Mikós-Singhi conjecture has been open for more than two decades. Only a few partial results of this conjecture are known so far. The most important one among them is that the conjecture holds for all n divisible by k. This claim can be proved directly by considering a random partition of our set of numbers into pairwise disjoint sets, each of size k, but it also follows immediately from Baranyai's partition theorem [9]. This theorem asserts that if $k \mid n$, then the family of all k-subsets of [n] can be partitioned into disjoint subfamilies so that each subfamily is a perfect k-matching. Since the total sum is nonnegative, among the n/k subsets from each subfamily, there must be at least one having a nonnegative sum. Hence there are no less than $\binom{n}{k}/(n/k) = \binom{n-1}{k-1}$ nonnegative k-sums in total. Besides this case, the conjecture is also known to be true for small k. It is not hard to check it for k = 2, and the case k = 3 was settled by Manickam [67], and by Marino and Chiaselotti [70] independently.

Let f(k) be the minimal number N such that $A(n,k) = \binom{n-1}{k-1}$ for all $n \ge N$. The Manickam-Miklós-Singhi conjecture states that $f(k) \le 4k$. The existence of such function fwas first demonstrated by Manickam and Miklós [68] by showing $f(k) \le (k-1)(k^k+k^2)+k$. Bhattacharya [10] found a new and shorter proof of existence of f later, but he didn't improve the previous bound. Very recently, Tyomkyn [101] obtained a better upper bound $f(k) \le k(4e \log k)^k \sim e^{ck \log \log k}$, which is still exponential.

In this chapter, we discuss a connection between the Manickam-Miklós-Singhi conjecture and a problem about matchings in dense uniform hypergraph. We call a hypergraph Hr-uniform if all the edges have size r. Denote by $\nu(H)$ the matching number of H, which is the maximum number of pairwise disjoint edges in H. For our application, we need the fact that if a (k - 1)-uniform hypergraph on n - 1 vertices has matching number at most n/k, then its number of edges cannot exceed $c\binom{n-1}{k-1}$ for some constant c < 1 independent of n, k. This is closely related to a special case of a long-standing open problem of Erdős [28], who in 1965 asked to determine the maximum possible number of edges of an r-uniform hypergraph H on n vertices with matching number $\nu(H)$. Erdős conjectured that the optimal case is when H is a clique or the complement of a clique, more precisely, for $\nu(H) < \lfloor n/r \rfloor$ the maximum possible number of edges is given by the following equation:

$$\max e(H) = \max\left\{ \binom{r[\nu(H)+1]-1}{r}, \ \binom{n}{r} - \binom{n-\nu(H)}{r} \right\}$$
(2.1.1)

For our application to the Manickam-Miklós-Singhi conjecture, it suffices to prove a weaker statement which bounds the number of edges as a function of the fractional matching number $\nu^*(H)$ instead of $\nu(H)$. To attack the latter problem we combine duality with a probabilistic technique together with an inequality by Feige [32] which bounds the probability that the sum of an arbitrary number of nonnegative independent random variables exceeds its expectation by a given amount. Using this machinery, we obtain the first polynomial upper bound $f(k) \leq 33k^2$, which substantially improves all the previous exponential estimates.

Theorem 2.1.2. Given integers n and k satisfying $n \ge 33k^2$, for any n real numbers $\{x_1, \dots, x_n\}$ whose sum is nonnegative, there are at least $\binom{n-1}{k-1}$ nonnegative k-sums.

Recall that earlier we mentioned the similarity between the Manickam-Miklós-Singhi conjecture and the Erdős-Ko-Rado theorem. When $n \ge 4k$, the conjectured extremal example is $x_1 = n - 1, x_2 = \cdots = x_n = -1$, where all the $\binom{n-1}{k-1}$ nonnegative k-sums use x_1 . For the Erdős-Ko-Rado theorem when n > 2k, the extremal family also consists of all the $\binom{n-1}{k-1}$ subsets containing one fixed element. It is a natural question to ask if this kind of structure is forbidden, can we obtain a significant improvement on the $\binom{n-1}{k-1}$ bound? A classical result of Hilton and Milner [45] asserts that if n > 2k and no element is contained in every k-subset, then the intersecting family has size at most $\binom{n-1}{k-1} - \binom{n-k-1}{k-1} + 1$, with the extremal example being one of the following two.

- Fix $x \in [n]$ and $X \subset [n] \setminus \{x\}$, |X| = k. The family \mathcal{F}_1 consists of X and all the k-subsets containing x and intersecting with X.
- Take $Y \subset [n]$, |Y| = 3. The family \mathcal{F}_2 consists of all the k-subsets of [n] which intersects Y with at least two elements.

It can be easily checked that both families are intersecting and $|\mathcal{F}_1| = \binom{n-1}{k-1} - \binom{n-k-1}{k-1} + 1$, $|\mathcal{F}_2| = 3\binom{n-3}{k-2} + \binom{n-3}{k-3}$. When k = 3, $|\mathcal{F}_1| = |\mathcal{F}_2|$ and their structures are non-isomorphic. For $k \ge 4$, $|\mathcal{F}_1| > |\mathcal{F}_2|$, so only the first construction is optimal.

Here we prove a Hilton-Milner type result about the minimum number of nonnegative k-sums. Call a number x_i large if its sum with any other k - 1 numbers x_j is nonnegative. We prove that if no x_i is large, then the $\binom{n-1}{k-1}$ bound can be greatly improved. We also show that there are two extremal structures, one of which is maximum for every k and the other only for k = 3. This result can be considered as an analogue of the two extremal cases mentioned above in the Hilton-Milner theorem.

Theorem 2.1.3. For any fixed integer k and sufficiently large n, and for any n real numbers x_1, \dots, x_n with $\sum_{i=1}^n x_i \ge 0$, where no x_i is large, the number N of different nonnegative k-sums is at least $\binom{n-1}{k-1} + \binom{n-k-1}{k-1} - 1$.

For large n, Theorem 2.1.3 (whose statement is tight) improves the $\binom{n-1}{k-1}$ bound in the nonnegative k-sum problem to $\binom{n-1}{k-1} + \binom{n-k-1}{k-1} - 1$ when large numbers are forbidden. This bound is asymptotically $(2 + o(1))\binom{n-1}{k-1}$.

Call a number x_i $(1 - \delta)$ -moderately large, if there are at least $(1 - \delta) \binom{n-1}{k-1}$ nonnegative k-sums using x_i , for some constant $0 \le \delta < 1$. In particular, when $\delta = 0$ this is the definition of a large number. If there is no $(1 - \delta)$ -moderately large number for some positive δ , we can prove a much stronger result asserting that at least a constant proportion of the $\binom{n}{k}$ k-sums are nonnegative. More precisely, we prove the following statement.

Theorem 2.1.4. There exists a positive function $g(\delta, k)$, such that for any fixed k and δ and all sufficiently large n, the following holds. For any set of n real numbers x_1, \dots, x_n with nonnegative sum in which no member is $(1 - \delta)$ -moderately large, the number N of nonnegative k-sums in the set is at least $g(\delta, k) {n \choose k}$.

The rest of this chapter is organized as follows. In the next section we present a quick proof of a slightly worse bound for the function f(k) defined above, namely, we show that $f(k) \leq 2k^3$. The proof uses a simple estimate on the number of edges in a hypergraph with a given matching number. The proof of Theorem 2.1.2 appears in Section 2.3, where we improve this estimate using more sophisticated probabilistic tools. In Section 2.4 we prove the Hilton-Milner type results Theorem 2.1.3 and 2.1.4. The last section contains some concluding remarks and open problems.

2.2 Nonnegative k-sums and hypergraph matchings

In this section we discuss the connection of the Manickam-Miklós-Singhi conjecture and hypergraph matchings, and verify the conjecture for $n \ge 2k^3$.

Without loss of generality, we can assume $\sum_{i=1}^{n} x_i = 0$ and $x_1 \ge x_2 \ge \cdots \ge x_n$ with $x_1 > 0$. If the sum of x_1 and the k-1 smallest numbers x_{n-k+2}, \cdots, x_n is nonnegative, there are already $\binom{n-1}{k-1}$ nonnegative k-sums by taking x_1 and any other k-1 numbers. Consequently we can further assume that $x_1 + x_{n-k+2} + \cdots + x_n < 0$. As all the numbers sum up to zero, we have

$$x_2 + \dots + x_{n-k+1} > 0 \tag{2.2.1}$$

Let *m* be the largest integer not exceeding n - k which is divisible by *k*, then $n - 2k + 1 \le m \le n - k$. Since the numbers are sorted in descending order, we have

$$x_2 + \dots + x_{m+1} \ge \frac{m}{n-k} (x_2 + \dots + x_{n-k+1}) > 0$$
 (2.2.2)

As mentioned in the introduction of this chapter, the Manickam-Mikós-Singhi conjecture holds when n is divisible by k by Baranyai's partition theorem, thus there are at least $\binom{m-1}{k-1} \ge \binom{n-2k}{k-1}$ nonnegative k-sums using only numbers from $\{x_2, \dots, x_{m+1}\}$. From now on we are focusing on counting the number of nonnegative k-sums involving x_1 . If this number plus $\binom{n-2k}{k-1}$ is at least $\binom{n-1}{k-1}$, then the Manickam-Miklós-Singhi conjecture is true.

Recall that in the proof of the case $k \mid n$, if we regard all the negative k-sums as edges in a k-uniform hypergraph, then the assumption that all numbers add up to zero provides us the fact that this hypergraph does not have a perfect k-matching. One can prove there are at least $\binom{n-1}{k-1}$ edges in the complement of such a hypergraph, which exactly tells the minimum number of nonnegative k-sums. We utilize the same idea to estimate the number of nonnegative k-sums involving x_1 . Construct a (k-1)-uniform hypergraph H on the vertex set $\{2, \dots, n\}$. The edge set E(H) consists of all the (k-1)-tuples $\{i_1, \dots, i_{k-1}\}$ corresponding to the negative k-sum $x_1 + x_{i_1} + \dots + x_{i_{k-1}} < 0$. Our goal is to show that e(H) = |E(H)| cannot be too large, and therefore there must be lots of nonnegative k-sums involving x_1 .

Denote by $\nu(H)$ the matching number of our hypergraph H, which is the maximum number of disjoint edges in H. By definition, every edge corresponds to a (k-1)-sum which is less than $-x_1$, thus the sum of the $(k-1)\nu(H)$ numbers corresponding to the vertices in the maximal matching is less than $-\nu(H)x_1$. On the other hand, all the remaining $n-1-(k-1)\nu(H)$ numbers are at most x_1 . Therefore $-x_1 = x_2 + \cdots + x_n \leq -\nu(H)x_1 + (n-1-(k-1)\nu(H))x_1$. By solving this inequality, we have the following lemma.

Lemma 2.2.1. The matching number $\nu(H)$ is at most n/k.

If the matching number of a hypergraph is known and n is large with respect to k, we are able to bound the number of its edges using the following lemma. We denote by \overline{H} the complement of the hypergraph H.

Lemma 2.2.2. If $n > k^3$, any (k-1)-uniform hypergraph H on n-1 vertices with matching number at most n/k has at least $\frac{1}{k+1} \binom{n-1}{k-1}$ edges missing from it.

Proof. Consider a random permutation σ on the n-1 vertices v_1, \dots, v_{n-1} of H. Let the random variables $Z_1 = 1$ if $(\sigma(v_1), \dots, \sigma(v_{k-1}))$ is an edge in H and 0 otherwise. Repeat the same process for the next k-1 indices and so on. Finally we will have at least $m \geq \frac{n-k}{k-1}$ random variables Z_1, \dots, Z_m . Let $Z = Z_1 + \dots + Z_m$. Z is always at most n/k since there is no matching of size larger than n/k. On the other hand, $\mathbb{E}Z_i$ is the probability that k-1 randomly chosen vertices form an edge in H, therefore $\mathbb{E}Z_i = e(H)/\binom{n-1}{k-1}$. Hence,

$$\frac{n}{k} \ge \mathbb{E}Z = m \frac{e(H)}{\binom{n-1}{k-1}} \ge \frac{n-k}{k-1} \frac{e(H)}{\binom{n-1}{k-1}}$$
(2.2.3)

The number of edges missing is equal to $e(\overline{H}) = \binom{n-1}{k-1} - e(H)$. By (2.2.3), $e(H) \leq (1 - \frac{1}{k}) \frac{n}{n-k} \binom{n-1}{k-1}$, therefore

$$e(\overline{H}) \geq \left[1 - \left(1 - \frac{1}{k}\right)\frac{n}{n-k}\right]\binom{n-1}{k-1}$$

$$\geq \left[1 - \left(1 - \frac{1}{k}\right)\frac{k^3}{k^3-k}\right]\binom{n-1}{k-1}$$

$$= \frac{1}{k+1}\binom{n-1}{k-1}$$
(2.2.4)

Now we can easily prove a polynomial upper bound for the function f(k) considered in Section 2.1, showing that $f(k) \leq 2k^3$.

Theorem 2.2.3. If $n \ge 2k^3$, then for any n real numbers $\{x_1, \dots, x_n\}$ whose sum is nonnegative, the number of nonnegative k-sums is at least $\binom{n-1}{k-1}$.

Proof. By Lemma 2.2.2, there are at least $\frac{1}{k+1} \binom{n-1}{k-1}$ edges missing in H, which also gives a lower bound for the number of nonnegative k-sums involving x_1 . Together with the previous $\binom{n-2k}{k-1}$ nonnegative k-sums without using x_1 , there are at least $\frac{1}{k+1} \binom{n-1}{k-1} + \binom{n-2k}{k-1}$ nonnegative k-sums in total. We claim that this number is greater than $\binom{n-1}{k-1}$ when $n \ge 2k^3$. This statement is equivalent to proving $\binom{n-2k}{k-1} / \binom{n-1}{k-1} \ge 1 - 1/(k+1)$, which can be completed as

follows:

$$\binom{n-2k}{k-1} / \binom{n-1}{k-1} = \left(1 - \frac{2k-1}{n-1}\right) \left(1 - \frac{2k-1}{n-2}\right) \cdots \left(1 - \frac{2k-1}{n-k+1}\right)$$

$$\ge 1 - \frac{(2k-1)(k-1)}{n-k+1}$$

$$\ge 1 - \frac{(2k-1)(k-1)}{2k^3 - k+1}$$

$$\ge 1 - \frac{1}{k+1}$$

$$(2.2.5)$$

The last inequality is because $(2k-1)(k-1)(k+1) = 2k^3 - k^2 - 2k + 1 \le 2k^3 - k + 1$. \Box

2.3 Fractional covers and small deviations

The method above verifies the Manickam-Miklós-Singhi conjecture for $n \ge 2k^3$ and improves the current best exponential lower bound $n \ge k(4e \log k)^k$ by Tyomkyn [101]. However if we look at Lemma 2.2.2 attentively, there is still some room to improve it. Recall our discussion of Erdős' conjecture in the introduction: if the conjecture is true in general, then in order to minimize the number of edges in a (k - 1)-hypergraph of a given matching number $\nu(H) = n/k$, the hypergraph must be either a clique of size (k - 1)(n/k + 1) - 1 or the complement of a clique of size n - n/k.

$$e(H) \sim \min\left\{ \binom{(1-1/k)n}{k-1}, \ \binom{n-1}{k-1} - \binom{n-n/k}{k-1} \right\} \sim (1-\frac{1}{k})^{k-1} \binom{n-1}{k-1} \le \frac{1}{2} \binom{n-1}{k-1}$$
(2.3.1)

In this case, the number of edges missing from H must be at least $\frac{1}{2} \binom{n-1}{k-1}$, which is far larger than the bound $\frac{1}{k+1} \binom{n-1}{k-1}$ in Lemma 2.2.2. If in our proof of Theorem 2.2.3, the coefficient before $\binom{n-1}{k-1}$ can be changed to a constant instead of the original $\frac{1}{k+1}$, the theorem can also be sharpened to $n \ge \Omega(k^2)$. Based on this idea, in the rest of this section we are going to prove Lemma 2.3.3, which asserts that $e(H) \ge c\binom{n-1}{k-1}$ for some constant c independent of n and k, and can be regarded as a strengthening of Lemma 2.2.2. Then we use it to prove our main result of this chapter, Theorem 2.3.5. In order to improve Lemma 2.2.2, instead of using the usual matching number $\nu(H)$, it suffices to consider its fractional relaxation, which is defined as follows.

$$\nu^*(H) = \max \sum_{e \in E(H)} w(e) \qquad w : E(H) \to [0, 1]$$

subject to $\sum_{i \in e} w(e) \le 1$ for every vertex *i*. (2.3.2)

Note that $\nu^*(H)$ is always greater or equal than $\nu(H)$. On the other hand, for our hypergraph we can prove the same upper bound for the fractional matching number $\nu^*(H)$ as in Lemma 2.2.1. Recall that H is the (k-1)-uniform hypergraph on the n-1 vertices $\{2, \dots, n\}$, whose edges are those (k-1)-tuples (i_1, \dots, i_{k-1}) corresponding to negative k-sums $x_1 + x_{i_1} + \dots + x_{i_k} < 0$.

Lemma 2.3.1. The fractional matching number $\nu^*(H)$ is at most n/k.

Proof. Choose a weight function $w : E(H) \to [0,1]$ which optimizes the linear program (2.3.2) and gives the fractional matching number $\nu^*(H)$, then $\nu^*(H) = \sum_{e \in E(H)} w(e)$. Two observations can be easily made: (i) if $e \in E(H)$, then $\sum_{i \in e} x_i < -x_1$; (ii) $x_i \leq x_1$ for any $i = 2, \dots, n$ since $\{x_i\}$ are in descending order. Therefore we can bound the fractional matching number in a few steps.

$$x_{1} + x_{2} + \dots + x_{n} = x_{1} + \sum_{i=2}^{n} \left(\sum_{i \in e} w(e) \right) x_{i} + \sum_{i=2}^{n} \left(1 - \sum_{i \in e} w(e) \right) x_{i}$$

$$\leq x_{1} + \sum_{e \in E(H)} \left(\sum_{i \in e} x_{i} \right) w(e) + \sum_{i=2}^{n} \left(1 - \sum_{i \in e} w(e) \right) x_{1}$$

$$\leq x_{1} + \sum_{e \in E(H)} w(e) (-x_{1}) + \left(n - 1 - \sum_{e \in E(H)} \sum_{i \in e} w(e) \right) x_{1}$$

$$= x_{1} - \nu^{*}(H) x_{1} + \left(n - 1 - (k - 1)\nu^{*}(H) \right) x_{1}$$

$$= (n - k\nu^{*}(H)) x_{1} \qquad (2.3.3)$$

Lemma 2.3.1 follows from this inequality and our assumption that $x_1 + \cdots + x_n = 0$ and $x_1 > 0$.

The determination of the fractional matching number is actually a linear programming problem. Therefore we can consider its dual problem, which gives the fractional covering number $\tau^*(H)$.

$$\tau^*(H) = \min \sum_i v(i) \qquad v: V(H) \to [0, 1]$$

subject to $\sum_{i \in e} v(i) \ge 1$ for every edge e . (2.3.4)

By duality we have $\tau^*(H) = \nu^*(H) \leq n/k$. Getting a upper bound for e(H) is equivalent to finding a function $v: V(H) \to [0, 1]$ satisfying $\sum_{i \in V(H)} v(i) \leq n/k$ that maximizes the number of (k-1)-tuples e where $\sum_{i \in e} v(i) \geq 1$. Since this number is monotone increasing in every v(i), we can assume that it is maximized by a function v with $\sum_{i \in V(H)} v(i) = n/k$.

The following lemma was established by Feige [32], and later improved by He, Zhang, and Zhang [44]. It bounds the tail probability of the sum of independent nonnegative random variables with given expectations. It is stronger than Markov's inequality in the sense that the number of variables m does not appear in the bound.

Lemma 2.3.2. Given m independent nonnegative random variables X_1, \dots, X_m , each of expectation at most 1, then

$$Pr\left(\sum_{i=1}^{m} X_i < m+\delta\right) \ge \min\left\{\delta/(1+\delta), \frac{1}{13}\right\}$$
(2.3.5)

Now we can show that the complement of the hypergraph H has at least constant edge density, which implies as a corollary that a constant proportion of the k-sums involving x_1 must be nonnegative.

Lemma 2.3.3. If $n \ge Ck^2$ with $C \ge 1$, and H is a (k-1)-uniform hypergraph on n-1 vertices with fractional covering number $\tau^*(H) = n/k$, then there are at least $\left(\frac{1}{13} - \frac{1}{2C}\right) \frac{(n-1)^{k-1}}{(k-1)!}$ (k-1)-sets which are not edges in H.

Proof. Choose a weight function $v: V(H) \to [0, 1]$ which optimizes the linear programming problem (2.3.4). Define a sequence of k-1 independent and identically distributed random variables X_1, \dots, X_{k-1} , such that for any $1 \leq j \leq k-1, 2 \leq i \leq n, X_j = v(i)$ with probability 1/(n-1). It is easy to compute the expectation of X_i , which is

$$\mathbb{E}X_i = \frac{1}{n-1} \sum_{i=2}^n v(i) = \frac{n}{k(n-1)}$$
(2.3.6)
Now we can estimate the number of (k-1)-tuples with sum less than 1. The probability of the event $\left\{\sum_{i=1}^{k-1} X_i < 1\right\}$ is basically the same as the probability that a random (k-1)tuple has sum less than 1, except that two random variables X_i and X_j might share a weight from the same vertex, which is not allowed for forming an edge. However, we assumed that n is much larger than k, so this error term is indeed negligible for our application. Note that for $i_1 < \cdots < i_{k-1}$, the probability that $X_j = v(i_j)$ for every $1 \le j \le k-1$ is equal to $1/(n-1)^{k-1}$.

$$e(\overline{H}) = \left| \left\{ i_{1} < \dots < i_{k-1} : v(i_{1}) + \dots + v(i_{k-1}) < 1 \right\} \right|$$

$$= \frac{(n-1)^{k-1}}{(k-1)!} \sum_{\text{distinct } i_{1},\dots,i_{k-1}} Pr\left[X_{1} = v(i_{1}),\dots,X_{k-1} = v(i_{k-1});\sum_{i=1}^{k-1} X_{i} < 1 \right]$$

$$\geq \frac{(n-1)^{k-1}}{(k-1)!} \left[Pr\left(\sum_{i=1}^{k-1} X_{i} < 1\right) - \sum_{l} \sum_{i \neq j} Pr\left(X_{i} = X_{j} = v(i_{l})\right) \right]$$

$$\geq \frac{(n-1)^{k-1}}{(k-1)!} \left[Pr\left(\sum_{i=1}^{k-1} X_{i} < 1\right) - \frac{\binom{k-1}{2}}{n-1} \right]$$

$$\geq \frac{(n-1)^{k-1}}{(k-1)!} \left[Pr\left(\sum_{i=1}^{k-1} X_{i} < 1\right) - \frac{1}{2C} \right]$$

$$(2.3.7)$$

The last inequality is because $n \ge Ck^2$ and $-1 \ge -3Ck + 2C$ for $C \ge 1, k \ge 1$, and the sum of these two inequalities implies that $\frac{(k-1)(k-2)}{2(n-1)} \le \frac{1}{2C}$.

Define $Y_i = X_i \cdot k(n-1)/n$ to normalize the expectations to $\mathbb{E}Y_i = 1$. Y_i 's are nonnegative because each vertex receives a nonnegative weight in the linear program (2.3.4). Applying Lemma 2.3.2 to Y_1, \dots, Y_{k-1} and setting m = k - 1, $\delta = (n - k)/n$, we get

$$Pr(X_{1} + \dots + X_{k-1} < 1) = Pr(Y_{1} + \dots + Y_{k-1} < k(n-1)/n)$$

$$\geq \min\left\{\frac{n-k}{2n-k}, \frac{1}{13}\right\}$$
(2.3.8)

When $n > Ck^2$ and $k \ge 2, C \ge 1$, we have

$$\frac{n-k}{2n-k} > \frac{Ck^2-k}{2Ck^2-k} = \frac{Ck-1}{2Ck-1} \ge \frac{1}{13}$$
(2.3.9)

Combining (2.3.7) and (2.3.8) we immediately obtain Lemma 2.3.3.

Corollary 2.3.4. If $n \ge Ck^2$ with $C \ge 1$, then there are at least $\left(\frac{1}{13} - \frac{1}{2C}\right)\frac{(n-1)^{k-1}}{(k-1)!}$ nonnegative k-sums involving x_1 .

Now we are ready to prove our main theorem of this chapter:

Theorem 2.3.5. If $n \ge 33k^2$, then for any n real numbers x_1, \dots, x_n with $\sum_{i=1}^n x_i \ge 0$, the number of nonnegative k-sums is at least $\binom{n-1}{k-1}$.

Proof. By the previous discussion, we know that there are at least $\binom{n-2k}{k-1}$ nonnegative k-sums using only x_2, \dots, x_n . By Corollary 2.3.4, there are at least $\left(\frac{1}{13} - \frac{1}{2 \cdot 33}\right) \frac{(n-1)^{k-1}}{(k-1)!} \geq \frac{2}{33} \frac{(n-1)^{k-1}}{(k-1)!}$ nonnegative k-sums involving x_1 . In order to prove the theorem, we only need to show that for $n \geq 33k^2$,

$$\frac{2}{33}\frac{(n-1)^{k-1}}{(k-1)!} + \binom{n-2k}{k-1} \ge \binom{n-1}{k-1}$$
(2.3.10)

Define an infinitely differentiable function $g(x) = \begin{pmatrix} x \\ k-1 \end{pmatrix} = \frac{x(x-1)\cdots(x-k+2)}{(k-1)!}$. It is not difficult to see g''(x) > 0 when x > k-1. Therefore

$$\binom{n-1}{k-1} - \binom{n-2k}{k-1} = g(n-1) - g(n-2k) \le [(n-1) - (n-2k)]g'(n-1) = (2k-1)g'(n-1)$$
(2.3.11)

$$g'(x) = \frac{(x-1)(x-2)\cdots(x-k+2)}{(k-1)!} + \frac{x(x-2)\cdots(x-k+2)}{(k-1)!} + \dots + \frac{x(x-1)\cdots(x-k+3)}{(k-1)!}$$

$$\leq (k-1)\frac{x(x-1)\cdots(x-k+3)}{(k-1)!}$$

$$\leq (k-1)\frac{x^{k-2}}{(k-1)!}$$
(2.3.12)

Combining (2.3.11) and (2.3.12),

$$\binom{n-1}{k-1} - \binom{n-2k}{k-1} \le (2k-1)g'(n-1) \le (2k-1)(k-1)\frac{(n-1)^{k-2}}{(k-1)!} \le \frac{2}{33}\frac{(n-1)^{k-1}}{(k-1)!}$$
(2.3.13)

The last inequality follows from our assumption $n \ge 33k^2$.

2.4 Hilton-Milner type results

In this section we prove two Hilton-Milner type results about the minimum number of nonnegative k-sums. The first theorem asserts that for sufficiently large n, if $\sum_{i=1}^{n} x_i \ge 0$ and no x_i is large, then there are at least $\binom{n-1}{k-1} + \binom{n-k-1}{k-1} - 1$ nonnegative k-sums.

Proof of Theorem 2.1.3. We again assume that $x_1 \ge \cdots \ge x_n$ and $\sum_{i=1}^n x_i$ is zero. Since x_1 is not large, we know that there exists a (k-1)-subset S_1 not containing 1, such that $x_1 + \sum_{i \in S_1} x_i < 0$. Suppose t is the largest integer so that there are t subsets S_1, \cdots, S_t , such that for any $1 \le j \le t$, S_j is disjoint from $\{1, \cdots, j\}$, has size at most j(k-1) and

$$x_1 + \dots + x_j + \sum_{i \in S_j} x_i < 0$$

As we explained above $t \ge 1$ and since $x_1 \ge \cdots \ge x_n$ we may also assume that S_j consists of the last $|S_j|$ indices in $\{1, \cdots, n\}$. By Corollary 2.3.4, for sufficiently large n, there are at least $\frac{1}{14} \binom{n-1}{k-1}$ nonnegative k-sums using x_1 . Note also that after deleting x_1 and $\{x_i\}_{i\in S_1}$, the sum of the remaining $n - 1 - |S_1| \ge n - k$ numbers is nonnegative. Therefore, again by Corollary 2.3.4, there are at least $\frac{1}{14} \binom{n-k-1}{k-1}$ nonnegative k-sums using x_2 but not x_1 . In the next step, we delete x_1, x_2 and $\{x_i\}_{i\in S_2}$ and bound the number of nonnegative k-sums involving x_3 but neither x_1 or x_2 by $\frac{1}{14} \binom{n-2k-1}{k-1}$. Repeating this process for 30 steps, we obtain

$$N \geq \frac{1}{14} \left[\binom{n-1}{k-1} + \binom{n-k-1}{k-1} + \dots + \binom{n-29k-1}{k-1} \right] \\> \frac{30}{14} \binom{n-29k-1}{k-1} > 2\binom{n-1}{k-1} > \binom{n-1}{k-1} + \binom{n-k-1}{k-1} - 1$$

where here we used the fact that $\frac{30}{14} > 2$ and n is sufficiently large (as a function of k).

If $2 \leq t < 30$, by the maximality of t, we know that the sum of x_{t+1} with any (k-1) numbers with indices not in $\{1, \dots, t+1\} \cup S_t$ is nonnegative. This gives us $\binom{n-(t+1)-|S_t|}{k-1} \geq \binom{n-tk-1}{k-1}$ nonnegative k-sums. We can also replace x_{t+1} by any x_i where

 $1 \leq i \leq t$ and the new k-sum is still nonnegative since $x_i \geq x_{t+1}$. Therefore,

$$N \ge (t+1)\binom{n-tk-1}{k-1} \ge (t+1)\binom{n-29k-1}{k-1} > 2\binom{n-1}{k-1}$$

for sufficiently large n. Thus the only remaining case is t = 1.

Recall that x_1 is not large, and hence $x_1 + (x_{n-k+2} + \cdots + x_n) < 0$. Suppose I is a (k-1)-subset of [2, n] such that $x_1 + \sum_{i \in I} x_i < 0$. If $2 \in I$, then $x_1 + x_2 + \sum_{i \in I \setminus \{2\}} x_i < 0$, this contradicts the assumption t = 1 since $|I \setminus \{2\}| = k - 2 \leq 2(k-1)$. Hence we can assume that all the (k-1)-subsets I_1, \cdots, I_m corresponding to negative k-sums involving x_1 belong to the interval [3, n]. Let N_1 be the number of nonnegative k-sums involving x_1 , and let N_2 be the number of nonnegative k-sums using x_2 but not x_1 , then

$$N \ge N_1 + N_2 = \left[\binom{n-1}{k-1} - m \right] + N_2.$$

In order to prove $N \ge \binom{n-1}{k-1} + \binom{n-k-1}{k-1} - 1$, we only need to establish the following inequality

$$N_2 \ge \binom{n-k-1}{k-1} + m - 1. \tag{2.4.1}$$

Observe that the subsets I_1, \dots, I_m satisfy some additional properties. First of all, if two sets I_i and I_j are disjoint, then by definition, $x_1 + \sum_{l \in I_i} x_l < 0$ and $x_2 + \sum_{l \in I_j} x_l \leq x_1 + \sum_{l \in I_j} x_l < 0$, summing them up gives $x_1 + x_2 + \sum_{l \in I_i \cup I_j} x_l < 0$ with $|I_i \cup I_j| = 2(k-1)$, which again contradicts the assumption t = 1. Therefore we might assume that $\{I_i\}_{1 \leq i \leq m}$ is an intersecting family. By the Erdős-Ko-Rado theorem,

$$m \le \binom{(n-2)-1}{(k-1)-1} = \binom{n-3}{k-2}.$$

The second observation is that if a (k-1)-subset $I \subset [3,n]$ is disjoint from some I_i , then $x_2 + \sum_{i \in I} x_i \ge 0$. Otherwise if $x_2 + \sum_{i \in I} x_i < 0$ and $x_1 + \sum_{k \in I_i} x_k < 0$, for the same reason this contradicts t = 1. Hence N_2 is bounded from below by the number of (k-1)-subsets $I \subset [3,n]$ such that I is disjoint from at least one of I_1, \dots, I_m . Equivalently we need to count the distinct (k-1)-subsets contained in some $J_i = [3,n] \setminus I_i$, all of which have sizes n-k-1. By the real version of the Kruskal-Katona theorem (see Ex.13.31(b) in [65]), if $m = \binom{x}{n-k-1}$ for some positive real number $x \ge n-k-1$, then $N_2 \ge \binom{x}{k-1}$. On the other

hand, it is already known that $1 \le m \le \binom{n-3}{k-2} = \binom{n-3}{n-k-1}$, thus $n-k-1 \le x \le n-3$. The only remaining step is to verify the following inequality for x in this range,

$$\binom{x}{k-1} \ge \binom{n-k-1}{k-1} + \binom{x}{n-k-1} - 1.$$
(2.4.2)

Let $f(x) = \binom{x}{k-1} - \binom{x}{n-k-1}$, note that when $x \le n-4 = (k-2) + (n-k-2)$,

$$f(x+1) - f(x) = \left[\begin{pmatrix} x+1\\k-1 \end{pmatrix} - \begin{pmatrix} x+1\\n-k-1 \end{pmatrix} \right] - \left[\begin{pmatrix} x\\k-1 \end{pmatrix} - \begin{pmatrix} x\\n-k-1 \end{pmatrix} \right]$$
$$= \begin{pmatrix} x\\k-2 \end{pmatrix} - \begin{pmatrix} x\\n-k-2 \end{pmatrix} \ge 0$$

The last inequality is because when n is large, $x \ge n - k - 1 > 2(k - 2)$. Moreover, $\binom{x}{t}$ is an increasing function for 0 < t < x/2, so when $x \le n - 4$, $\binom{x}{n-k-2} = \binom{x}{x-(n-k-2)} \le \binom{x}{k-2}$.

Therefore we only need to verify (2.4.2) for $n - k - 1 \le x < n - k$, which corresponds to $1 \le m \le n - k - 1$. For m = 1, (2.4.2) is obvious, so it suffices to look at the case $m \ge 2$. The number of distinct (k - 1)-subsets of J_1 or J_2 is minimized when $|J_1 \cap J_2| = n - k - 2$, which, by the inclusion-exclusion principle, gives

$$N_2 \ge 2\binom{n-k-1}{k-1} - \binom{n-k-2}{k-1} = \binom{n-k-1}{k-1} + \binom{n-k-2}{k-2}.$$

So (2.4.1) is also true for $2 \le m \le \binom{n-k-2}{k-2} + 1$. It is easy to see that for $k \ge 3$ and n sufficiently large, $n-k-1 \le \binom{n-k-2}{k-2} + 1$. For k = 2, we have x = n-3 and (2.4.2) becomes $\binom{n-3}{1} \ge \binom{n-3}{1} + \binom{n-3}{n-3} - 1$, which is true and completes the proof.

Remark 1. In order for all the inequalities to be correct, we only need $n > Ck^2$. By carefully analyzing the above computations, one can check that C = 500 is enough.

Remark 2. Note that in the proof, the equality (2.4.1) holds in two different cases. The first case is when m = 1, which means $x_1 + x_{n-k+2} + \cdots + x_n < 0$ but any other k-sums involving x_1 are nonnegative. All the other nonnegative k-sums are formed by x_2 together with any (k-1)-subsets not containing x_{n-k+2}, \cdots, x_n . This case is realizable by the following construction: $x_1 = k(k-1)n, x_2 = n-2, x_3 = \cdots = x_{n-k+1} = -1, x_{n-k+2} = \cdots = x_n = -(kn+1)$. The second case is in (2.4.2) when x = n - 4 and x = n - k - 1 holds simultaneously, which gives k = 3. In this case, $m = \binom{n-3}{n-4} = n - 3$, and the Kruskal-Katona theorem holds with

equality for the (n-4)-subsets J_1, \dots, J_{n-3} . That is to say, the negative 3-sums using x_1 are $x_1 + x_i + x_n$ for $3 \le i \le n-1$, while the nonnegative 3-sums containing x_2 but not x_1 are $x_2 + x_i + x_j$ for $3 \le i < j \le n-1$. This case can also be achieved by setting $x_1 = x_2 = 1$, $x_3 = \dots = x_{n-1} = \frac{1}{2(n-3)}$, and $x_n = -\frac{3}{2}$. For large n, these are the only two possible configurations achieving equality in Theorem 2.1.3.

Next we prove Theorem 2.1.4, which states that if $\sum_i x_i \ge 0$ and no x_i is moderately large, then at least a constant proportion of the $\binom{n}{k}$ k-sums are nonnegative.

Proof of Theorem 2.1.4. Suppose t is the largest integer so that there are t subsets S_1, \dots, S_t such that for any $1 \leq j \leq t$, S_j is disjoint from $\{1, \dots, j\}$, has at most j(k-1) elements, and

$$x_1 + \dots + x_j + \sum_{i \in S_j} x_i < 0.$$

By the maximality of t, the sum of x_{t+1} and any k-1 numbers x_i with indices from $[n] \setminus (\{1, \dots, t+1\} \cup S_t)$ is nonnegative, so there are at least $\binom{n-tk-1}{k-1}$ nonnegative k-sums using x_{t+1} . Since x_{t+1} is not $(1-\delta)$ -moderately large,

$$\binom{n-tk-1}{k-1} < (1-\delta)\binom{n-1}{k-1}.$$

For sufficiently large n, this is asymptotically equivalent to

$$\left(1 - \frac{tk}{n}\right)^{k-1} < 1 - \delta.$$

Since

$$\left(1-\frac{tk}{n}\right)^{k-1} > 1-\frac{tk(k-1)}{n},$$

we have

$$t > \frac{n}{k^2}\delta$$

Recall that by Corollary 2.3.4, for each $i = 1, \dots, \frac{n}{k^2}\delta$, x_i gives at least $\frac{1}{14}\binom{n-(i-1)k-1}{k-1}$ non-

negative k-sums, therefore

$$N \geq \frac{1}{14} \left[\binom{n-1}{k-1} + \dots + \binom{n - (\frac{n}{k^2}\delta)k - 1}{k-1} \right]$$

$$\geq \frac{n\delta}{14k^2} \binom{n - (\frac{n}{k^2}\delta)k - 1}{k-1}$$

$$= \frac{\delta}{14k} \binom{n}{k} \left(1 - \frac{\delta n/k}{n-1} \right) \cdots \left(1 - \frac{\delta n/k}{n-k+1} \right)$$

$$\geq \frac{\delta}{14k} \binom{n}{k} \left(1 - \frac{\delta n}{n-k+1} \right)$$

Since $\delta < 1$, when $n \ge \frac{k-1}{1-\sqrt{\delta}}$, we have $\frac{\delta n}{n-k+1} \le \sqrt{\delta}$. Therefore setting $g(\delta,k) = \frac{\delta(1-\sqrt{\delta})}{14k}$ completes the proof.

2.5 Concluding remarks

In this chapter, we have proved that if $n > 33k^2$, any n real numbers with a nonnegative sum have at least $\binom{n-1}{k-1}$ nonnegative k-sums, thereby verifying the Manickam-Miklós-Singhi conjecture in this range. Because of the inequality $\binom{n-2k}{k} + C\binom{n-1}{k-1} \ge \binom{n-1}{k-1}$ we used, our method will not give a better range than the quadratic one, and we did not try hard to compute the best constant in the quadratic bound. It would be interesting to decide if the Manickam-Miklós-Singhi conjecture can be verified for a linear range n > ck. Perhaps some algebraic methods or structural analysis of the extremal configurations will help.

Feige [32] conjectures that the constant 1/13 in Lemma 2.3.2 can be improved to 1/e. This is a special case of a more general question suggested by Samuels [86]. He asked to determine, for a fixed m, the infimum of $Pr(X_1 + \cdots + X_k < m)$, where the infimum is taken over all possible collections of k independent nonnegative random variables X_1, \cdots, X_k with given expectations μ_1, \cdots, μ_k . For k = 1 the answer is given by Markov's inequality. Samuels [86, 87] solved this question for $k \leq 4$, but for all $k \geq 5$ his problem is still completely open.

As pointed out to us by Andrzej Ruciński, part of our reasoning in Section 2.3 implies that the function A(n, k) defined in the first page is precisely $\binom{n}{k}$ minus the maximum possible number of edges in a k-uniform hypergraph on n vertices with fractional covering number strictly smaller than n/k. Indeed, given n reals x_1, \ldots, x_n with sum zero and only A(n,k)nonnegative k-sums, we may assume that the absolute value of each x_i is smaller than 1/k(otherwise simply multiply all of them by a sufficiently small positive real.) Next, add a sufficiently small positive ϵ to each x_i , keeping each x_i smaller than 1/k and keeping the sum of any negative k-tuple below zero (this is clearly possible.) Note that the sum of these new reals, call them x'_i , is strictly positive and the number of positive k-sums is A(n,k). Put $\nu(i) = 1/k - x_i'$ and observe that $\sum_i \nu(i) < n/k$ and the k -uniform hypergraph whose edges are all k-sets e for which $\sum_{i \in e} \nu(i) \ge 1$ has exactly $\binom{n}{k} - A(n,k)$ edges. Therefore, there is a k-uniform hypergraph on n vertices with fractional covering number strictly smaller than n/k and at least $\binom{n}{k} - A(n,k)$ edges. Conversely, given a k-uniform hypergraph H on n vertices and a fractional covering of it $\nu: V(H) \mapsto [0,1]$ with $\sum_i \nu(i) = n/k - \delta < n/k$ and $\sum_{i \in e} \nu(i) \ge 1$ for each $e \in E(H)$, one can define $x_i = \frac{1}{k} - \frac{\delta}{n} - \nu(i)$ to get a set of n reals whose sum is zero, in which the number of nonnegative k-sums is at most $\binom{n}{k} - |E(H)|$ (as the sum of the numbers x_i for every k-set forming an edge of H is at most $1 - \frac{k\delta}{n} - 1 < 0$). This implies the desired equality, showing that the problem of determining A(n,k) is equivalent to that of finding the maximum possible number of edges of a k-uniform hypergraph on n vertices with fractional covering number strictly smaller than n/k. Note that this is equivalent to the problem of settling the fractional version of the conjecture of Erdős for the extremal case of fractional matching number < n/k.

Acknowledgment The author would like to thank Andrzej Ruciński for helpful discussions and comments, and Nati Linial for inspiring conversations and intriguing questions which led us to the results in Section 2.4.

CHAPTER 3

The size of hypergraph and its matching number

3.1 Introduction

A k-uniform hypergraph, or simply k-graph, is a pair H = (V, E), where V = V(H) is a finite set of vertices, and $E = E(H) \subseteq {V \choose k}$ is a family of k-element subsets of V called edges. A matching in H is a set of disjoint edges in E(H). We denote by $\nu(H)$ the size of the largest matching, i.e., the maximum number of disjoint edges in H. The problem of finding the maximum matching in a hypergraph has many applications in various different areas of mathematics, computer science, and even computational chemistry. Yet although the graph matching problem is fairly well-understood, and solvable in polynomial time, most of the problems related to hypergraph matching tend to be very difficult and remain unsolved. Indeed, the hypergraph matching problem is known to be NP-hard even for 3uniform hypergraphs, without any good approximation algorithm.

One of the most basic open questions in this area was raised in 1965 by Erdős [28], which was briefly mentioned in the last chapter. Erdős asked to determine the maximum possible number of edges that can appear in any k-uniform hypergraph with matching number $\nu(H) < t \leq \frac{n}{k}$ (equivalently, without any t pairwise disjoint edges). He conjectured that this problem has only two extremal constructions. The first one is a clique consisting of all the k-subsets on kt - 1 vertices, which obviously has matching number t - 1. The second example is a k-uniform hypergraph on n vertices containing all the edges intersecting a fixed set of t - 1 vertices, which also forces the matching number to be at most t - 1. Neither construction is uniformly better than the other across the entire parameter space, so the conjectured bound is the maximum of these two possibilities. Note that in the second case, the complement of this hypergraph is a clique on n-t+1 vertices together with t-1 isolated vertices, and thus the original hypergraph has $\binom{n}{k} - \binom{n-t+1}{k}$ edges.

Conjecture 3.1.1. Every k-uniform hypergraph H on n vertices with matching number $\nu(H) < t \leq \frac{n}{k}$ satisfies

$$e(H) \le \max\left\{ \binom{kt-1}{k}, \binom{n}{k} - \binom{n-t+1}{k} \right\}.$$
(3.1.1)

In addition to being important in its own right, this Erdős conjecture has several interesting applications, which we will discuss in the concluding remarks of this chapter. Yet although it is more than forty years old, only partial results have been discovered so far. In the case t = 2, the condition simplifies to the requirement that every pair of edges intersects, so Conjecture 3.1.1 is thus equivalent to a classical theorem of Erdős, Ko, and Rado [30]: that any intersecting family of k-subsets on $n \ge 2k$ elements has size at most $\binom{n-1}{k-1}$. The graph case (k = 2) was separately verified in [29] by Erdős and Gallai. For general fixed t and k, Erdős [28] proved his conjecture for sufficiently large n. Frankl [34] showed that Conjecture 3.1.1 was asymptotically true for all n by proving the weaker bound $e(H) \le (t-1)\binom{n-1}{k-1}$.

A short calculation shows that when $t \leq \frac{n}{k+1}$, we always have $\binom{n}{k} - \binom{n-t+1}{k} > \binom{kt-1}{k}$, so the potential extremal example in this case has all edges intersecting a fixed set of t-1vertices. One natural question is then to determine the range of t (with respect to n and $k \geq 3$) for which the maximum is indeed equal to $\binom{n}{k} - \binom{n-t+1}{k}$, i.e., where the second case is optimal. Recently, Frankl, Rödl, and Ruciński [36] studied 3-uniform hypergraphs (k = 3), and proved that for $t \leq n/4$, the maximum was indeed $\binom{n}{3} - \binom{n-t+1}{3}$, establishing the conjecture in that range. For general $k \geq 4$, Bollobás, Daykin, and Erdős [14] explicitly computed the bounds achieved by the proof in [28], showing that the conjecture holds for $t < \frac{n}{2k^3}$. Frankl and Furëdi [34] established the result in a different range $t < (\frac{n}{100k})^{1/2}$, which improves the original bound when k is large relative to n. In this chapter, we extend the range in which the Erdős conjecture holds to all $t < \frac{n}{3k^2}$.

Theorem 3.1.2. For any integers n, k, t satisfying $t < \frac{n}{3k^2}$, every k-uniform hypergraph on n vertices without t disjoint edges contains at most $\binom{n}{k} - \binom{n-t+1}{k}$ edges.

To describe the improvement, we first outline the Erdős proof for the case $t < \frac{n}{2k^3}$. Let v be a vertex of highest degree. By induction on t we get t-1 disjoint edges F_1, \dots, F_{t-1} not containing v. If deg(v) is larger than $k(t-1)\binom{n-2}{k-2}$, which is the maximum possible number of edges meeting the k(t-1) vertices in $\bigcup_{i=1}^{t-1} F_i$, then we can find t disjoint edges. Otherwise, the number of edges meeting any of F_i is at most $k(t-1) \cdot k(t-1)\binom{n-2}{k-2}$, which turns out to be less than the total number of edges provided $n \ge 2k^3 t$. We observe that in the first part of the argument, we are also done if the t-th largest degree is greater than $2t\binom{n-2}{k-2}$. This puts a tighter constraint on the sum of the degrees of the k(t-1) vertices in $\bigcup_{i=1}^{t-1} F_i$, allowing the second part to go through under the relaxed assumption $n \ge 3k^2t$. The fact that t vertices of less large degrees are enough to find t disjoint edges leads naturally to the following multicolored version of the Erdős conjecture, which was also considered independently by Aharoni and Howard [1],

Conjecture 3.1.3. Let $\mathcal{F}_1, \ldots, \mathcal{F}_t$ be families of subsets in $\binom{[n]}{k}$. If $|\mathcal{F}_i| > \max\left\{\binom{n}{k} - \binom{n-t+1}{k}, \binom{kt-1}{k}\right\}$ for all $1 \le i \le t$, then there is a "rainbow" matching of size t: one that contains exactly one edge from each family.

The k = 2 case of this conjecture was established by Meshulam (see [1]). To obtain Theorem 3.1.2, we prove an asymptotic version of Conjecture 3.1.3, by showing that a rainbow matching exists whenever $|\mathcal{F}_i| > (t-1) \binom{n-1}{k-1}$ for every $1 \le i \le t$.

The rest of this chapter is organized as follows. In the next section, we describe the so-called *shifting* method, which is a well known technique in extremal set theory, and use it to prove some preliminary results. In Section 3.3 we first prove the multicolored Erdős conjecture asymptotically, and then use it to prove Theorem 3.1.2. There, we also use the same argument to show that Conjecture 3.1.3 holds for all $t < \frac{n}{3k^2}$. The last section contains some concluding remarks and open problems.

3.2 Shifting

In extremal set theory, one of the most important and widely-used tools is the technique of shifting, which allows us to limit our attention to sets with certain structure. In this section we will only state and prove the relevant results for Section 3.3. For more background on the applications of shifting in extremal set theory, we refer the reader to the survey [34] by Frankl.

Given a family \mathcal{F} of equal-size subsets of [n], for integers $1 \leq j < i \leq n$, we define the (i, j)-shift map S_{ij} as follows: for any set $F \in \mathcal{F}$,

$$S_{ij}(F) = \begin{cases} F \setminus \{i\} \cup \{j\}, & \text{iff } i \in F, \ j \notin F \text{ and } F \setminus \{i\} \cup \{j\} \notin \mathcal{F}; \\ F, & \text{otherwise.} \end{cases}$$

Also, we denote the family after shifting as

$$S_{ij}(\mathcal{F}) = \{S_{ij}(F) : F \in \mathcal{F}\} .$$

Lemma 3.2.1. The shift map S_{ij} satisfies the following properties.

- (*i*) $|S_{ij}(\mathcal{F})| = |\mathcal{F}|.$
- (ii) If \mathcal{F} is k-uniform, then so is $S_{ij}(\mathcal{F})$.
- (iii) If the families $\mathcal{F}_1, \ldots, \mathcal{F}_t$ have the property that no subsets $F_1 \in \mathcal{F}_1, \ldots, F_t \in \mathcal{F}_t$ are pairwise disjoint, then the shifted families $S_{ij}(\mathcal{F}_1), \ldots, S_{ij}(\mathcal{F}_t)$ still preserve this property.

Proof. Claims (i) and (ii) are obvious. For (iii), assume that the statement is false, i.e., we have $F_i \in \mathcal{F}_i$ such that $S_{ij}(F_1), \ldots, S_{ij}(F_t)$ are pairwise disjoint, while F_1, \ldots, F_t are not. Without loss of generality, $F_1 \cap F_2 \neq \emptyset$. Next, observe that whenever $S_{ij}(F_k) \neq F_k$, we also have $j \in S_{ij}(F_k)$, so the pairwise disjointness of the $S_{ij}(F_k)$ implies that the only possible case (re-indexing if necessary) is for $S_{ij}(F_1) = F_1 \setminus \{i\} \cup \{j\}$, and $S_{ij}(F_k) = F_k$ for every $k \geq 2$. Note also that since F_1 and F_2 intersect while $S_{ij}(F_1)$ and $S_{ij}(F_2)$ do not, we must have $i \in F_2$ and $j \notin F_2$.

Therefore the only reason that $S_{ij}(F_2) = F_2$ is because $F'_2 = F_2 \setminus \{i\} \cup \{j\}$ is already in \mathcal{F}_2 . The pair of disjoint sets $S_{ij}(F_1)$ and $S_{ij}(F_2) = F_2$ have the same union as the pair of disjoint sets F_1 and F'_2 . Using the pairwise disjointness of the $S_{ij}(F_k)$, we conclude that the sets $F_1, F'_2, F_3, \ldots, F_t$ are pairwise disjoint as well, contradicting our initial assumption. \Box

In practice, we often combine the shifting technique with induction on the number of elements in the underlying set. Indeed, let us apply the shifts $\{S_{ni}\}_{1 \leq i \leq n-1}$ successively, and with slight abuse of notation, let us again call the resulting families $\mathcal{F}_1, \ldots, \mathcal{F}_t$. Create from each \mathcal{F}_i two sub-families based on containment of the final element n:

$$\mathcal{F}_i(n) = \{F \setminus \{n\} : F \in \mathcal{F}_i, n \in F\} ,$$
$$\mathcal{F}_i(\bar{n}) = \{F \qquad : F \in \mathcal{F}_i, n \notin F\} .$$

It turns out that the rainbow matching number does not increase by this decomposition.

Lemma 3.2.2. Let $\mathcal{F}_1, \ldots, \mathcal{F}_t$ be the shifted families, where each \mathcal{F}_i is k_i -uniform and $\sum_{i=1}^t k_i \leq n$. Suppose that no subsets $F_1 \in \mathcal{F}_1, \ldots, F_t \in \mathcal{F}_t$ are pairwise disjoint. Then, for any $0 \leq s \leq t$, the families $\mathcal{F}_1(n), \ldots, \mathcal{F}_s(n), \mathcal{F}_{s+1}(\bar{n}), \ldots, \mathcal{F}_t(\bar{n})$ still have the same property.

Proof. Assume for the sake of contradiction that there exist pairwise disjoint sets $F_1 \in \mathcal{F}_1(n), \ldots, F_s \in \mathcal{F}_s(n), F_{s+1} \in \mathcal{F}_{s+1}(\bar{n}), \ldots, F_t \in \mathcal{F}_t(\bar{n})$. By definition of $\mathcal{F}_i(n)$ and $\mathcal{F}_i(\bar{n})$, we know that $F_i \cup \{n\} \in \mathcal{F}_i$ for $1 \le i \le s$, and $F_i \in \mathcal{F}_i$ for $s+1 \le i \le t$. The size of $\bigcup_{i=1}^t F_i$ is equal to

$$\sum_{i=1}^{t} |F_i| = \sum_{i=1}^{s} (k_i - 1) + \sum_{i=s+1}^{t} k_i = \sum_{i=1}^{t} k_i - s \le n - s,$$

so there exist distinct elements $x_1, \ldots, x_s \notin \bigcup_{i=1}^t F_i$. Since $F_i \cup \{n\}$ is invariant under the shift S_{nx_i} , the set $F_i \cup \{x_i\} = (F_i \cup \{n\}) \setminus \{n\} \cup \{x_i\}$ must also be in the family \mathcal{F}_i . Taking $F'_i = F_i \cup \{x_i\}$ for $1 \leq i \leq s$, together with F_i for $s + 1 \leq i \leq t$, it is clear that we have found pairwise disjoint sets from \mathcal{F}_i , contradiction. \Box

3.3 Main result

In this section, we discuss the Erdős conjecture and its multicolored generalizations, and prove the original conjecture for the range $t < \frac{n}{3k^2}$. The colored interpretation arises from considering the collection of families \mathcal{F}_i as a single uniform hypergraph (possibly with repeated edges) on the vertex set [n], where each set in \mathcal{F}_i introduces a hyperedge colored in the *i*-th color. The following lemma is a multicolored generalization of Theorem 10.3 in [34], and provides a sufficient condition for a multicolored hypergraph to contain a rainbow matching of size t.

Lemma 3.3.1. Let $\mathcal{F}_1, \ldots, \mathcal{F}_t$ be families of subsets of [n] such that for each i, \mathcal{F}_i only contains sets of size $k_i, |\mathcal{F}_i| > (t-1) \binom{n-1}{k_i-1}$, and $n \ge \sum_{i=1}^t k_i$. Then there exist t pairwise disjoint sets $F_1 \in \mathcal{F}_1, \ldots, F_t \in \mathcal{F}_t$.

Proof. We proceed by induction on t and n. The case t = 1 is trivial. For general t, we can also handle all minimal cases of the form $n = \sum_{i=1}^{t} k_i$. Indeed, consider a uniformly random permutation π of [n], and define a series of indicator random variables $\{X_i\}$ as follows: $X_1 = 1$ iff $\{\pi(1), \ldots, \pi(k_1)\}$ is a set in \mathcal{F}_1 and $X_1 = 0$ otherwise, and in general, $X_j = 1$ iff $\{\pi(k_1 + \cdots + k_{j-1} + 1), \ldots, \pi(k_1 + \cdots + k_j)\}$ is a set in \mathcal{F}_j . We assume that there are no t disjoint sets from different families, so we deterministically have:

$$X_1 + \dots + X_t \le t - 1. \tag{3.3.1}$$

On the other hand, it is easy to see that the expectation of X_i is the probability that a random k_i -set is in \mathcal{F}_i , so

$$\mathbb{E}X_i = \frac{|\mathcal{F}_i|}{\binom{n}{k_i}}.$$

Yet we know that for every *i*, we have $|\mathcal{F}_i| > (t-1) \binom{n-1}{k_i-1}$, so

$$\mathbb{E}X_i > \frac{(t-1)\binom{n-1}{k_i-1}}{\binom{n}{k_i}} = (t-1)\frac{k_i}{n}$$

Summing these inequalities over $1 \le i \le t$, we obtain that $\sum_{i=1}^{t} \mathbb{E}X_i > t-1$, a contradiction to (3.3.1).

Now we consider a generic instance with $n > \sum_{i=1}^{t} k_i$, and inductively assume that all instances with smaller n are known. By Lemma 3.2.1, after applying all shifts $\{S_{ni}\}_{1 \le i \le n-1}$, we obtain families in which any rainbow t-matching can be pulled back to a rainbow tmatching in $\{\mathcal{F}_i\}$. For convenience we still call the shifted families $\{\mathcal{F}_i\}$. Our next step is to partition each \mathcal{F}_i into $\mathcal{F}_i(n) \cup \mathcal{F}_i(\bar{n})$, but in order to avoid empty sets, we first dispose of the case when there is some $k_i = 1$ with $\{n\} \in \mathcal{F}_i$. After re-indexing, we may assume that this is \mathcal{F}_1 . Since $|\mathcal{F}_i| > (t-1) {n-1 \choose k_i-1}$ and there are at most ${n-1 \choose k_i-1}$ sets containing n, every other \mathcal{F}_i has more than $(t-2) {n-1 \choose k_i-1}$ sets which in fact lie in [n-1]. By induction on the t-1 sizes k_2, \ldots, k_t , we find t-1 such disjoint sets from $\mathcal{F}_2, \ldots, \mathcal{F}_t$ which, together with $\{n\} \in \mathcal{F}_1$, establish the claim.

Returning to the general case, since $|\mathcal{F}_i| = |\mathcal{F}_i(n)| + |\mathcal{F}_i(\bar{n})|$ and our size condition is

$$|\mathcal{F}_i| > (t-1)\binom{n-1}{k_i-1} = (t-1)\binom{n-2}{k_i-2} + (t-1)\binom{n-2}{k_i-1},$$

we conclude that for each *i*, either $|\mathcal{F}_i(n)| > (t-1)\binom{n-2}{k_i-2}$ or $|\mathcal{F}_i(\bar{n})| > (t-1)\binom{n-2}{k_i-1}$. Without loss of generality, we may assume that $|\mathcal{F}_i(n)| > (t-1)\binom{n-2}{k_i-2}$ for $1 \le i \le s$, and $|\mathcal{F}_i(\bar{n})| > (t-1)\binom{n-2}{k_i-1}$ for $s+1 \le i \le t$. Note that \mathcal{F}_i is (k_i-1) -uniform for $1 \le i \le s$ and k_i -uniform for $s+1 \le i \le t$, and the base set now has n-1 elements. Induction on *n* and Lemma 3.2.2 then produce *t* disjoint sets from different families.

As mentioned in the introduction, the conjectured extremal hypergraph when $t \leq \frac{n}{k+1}$ is the hypergraph consisting of all edges intersecting a fixed set of size t-1. If we inspect the vertex degree sequence of this hypergraph, we observe that although there are t-1 vertices with high degree $\binom{n-1}{k-1}$, the remaining vertices only have degree $\binom{n-1}{k-1} - \binom{n-t}{k-1}$. For small t, this is asymptotically about $(t-1)\binom{n-2}{k-2}$, which is much smaller than $\binom{n-1}{k-1} = \frac{n-1}{k-1}\binom{n-2}{k-2}$. The following corollary of Lemma 3.3.1 shows that this sort of phenomenon generally occurs when hypergraphs satisfy the conditions in the Erdős conjecture.

Corollary 3.3.2. If a k-uniform hypergraph H on n vertices has t distinct vertices v_1, \ldots, v_t with degrees $d(v_i) > 2(t-1) \binom{n-2}{k-2}$, and $kt \leq n$, then H contains t disjoint edges.

Proof. Let H_i be a (k - 1)-uniform hypergraph containing all the subsets of

 $V(H) \setminus \{v_1, \ldots, v_t\}$ of size k - 1 which together with v_i form an edge of H. For any fixed $1 \leq i \leq t$ and $j \neq i$, there are at most $\binom{n-2}{k-2}$ edges of H containing both vertices v_i and v_j . Therefore for every hypergraph H_i ,

$$e(H_i) \ge d(v_i) - (t-1)\binom{n-2}{k-2} > (t-1)\binom{n-2}{k-2} \ge (t-1)\binom{n-t-1}{k-2}$$

Since every hypergraph H_i is (k-1)-uniform and has n-t vertices, we can use Lemma 3.3.1 with $\mathcal{F}_i = E(H_i)$, $k_i = k - 1$ and n replaced by n - t, to find t disjoint edges $e_1 \in E(H_1)$, $\ldots, e_t \in E(H_t)$. Taking the edges $e_i \cup \{v_i\} \in E(H)$, we obtain t disjoint edges in the original hypergraph H.

Now we are ready to prove our main result, Theorem 3.1.2, which states that for $t < \frac{n}{3k^2}$, every k-uniform hypergraph on n vertices without t disjoint edges contains at most $\binom{n}{k} - \binom{n-t+1}{k}$ edges.

Proof of Theorem 3.1.2: We proceed by induction on t. The base case t = 1 is trivial, so we consider the general case, assuming that the t - 1 case is known. Suppose $e(H) > \binom{n}{k} - \binom{n-t+1}{k}$, and let us seek t disjoint edges in H. We first consider the situation when there is a vertex v of degree $d(v) > k(t-1)\binom{n-2}{k-2}$. Let H_v be the sub-hypergraph induced by the vertex set $V(H) \setminus \{v\}$. Since there are at most $\binom{n-1}{k-1}$ edges containing v,

$$e(H_v) \ge e(H) - \binom{n-1}{k-1} > \binom{n}{k} - \binom{n-t+1}{k} - \binom{n-1}{k-1} = \binom{n-1}{k} - \binom{(n-1)-(t-1)+1}{k}$$

By induction, there are t - 1 disjoint edges e_1, \ldots, e_{t-1} in H_v , spanning (t - 1)k distinct vertices $u_1, \ldots, u_{(t-1)k}$. Note that the number of edges containing v and any vertex u_j is at most $\binom{n-2}{k-2}$. Therefore since we assumed that $d(v) > k(t-1)\binom{n-2}{k-2}$, there must be another edge e_t which contains v but avoids $u_1, \ldots, u_{(t-1)k}$. We then have t disjoint edges e_1, \ldots, e_t in H.

Now suppose that the maximum vertex degree in H is at most $k(t-1)\binom{n-2}{k-2}$. After re-indexing the vertices, we may assume that $k(t-1)\binom{n-2}{k-2} \ge d(v_1) \ge \cdots \ge d(v_n)$. If the *t*-th largest degree satisfies $d(v_t) > 2(t-1)\binom{n-2}{k-2}$, then Corollary 3.3.2 immediately produces *t* disjoint edges in H, so we may also assume for the remainder that $d(v_t) \le 2(t-1)\binom{n-2}{k-2}$. By induction (with room to spare), we also know that there are t-1 disjoint edges in H, spanning (t-1)k vertices. Among these vertices, the t-1 largest degrees are at most $k(t-1)\binom{n-2}{k-2}$ by our maximum degree assumption, while the remaining (t-1)(k-1) vertices cannot have degrees exceeding $d(v_t) \leq 2(t-1)\binom{n-2}{k-2}$. Therefore the sum of degrees of these (t-1)k vertices is at most

$$(t-1) \cdot k(t-1) \binom{n-2}{k-2} + (t-1)(k-1) \cdot 2(t-1) \binom{n-2}{k-2} = (t-1)^2 (3k-2) \binom{n-2}{k-2} + (t-1)(k-1) \cdot 2(t-1) \binom{n-2}{k-2} = (t-1)^2 (3k-2) \binom{n-2}{k-2} + (t-1)(k-1) \cdot 2(t-1) \binom{n-2}{k-2} = (t-1)^2 (3k-2) \binom{n-2}{k-2} + (t-1)(k-1) \cdot 2(t-1) \binom{n-2}{k-2} = (t-1)^2 (3k-2) \binom{n-2}{k-2} + (t-1)(k-1) \cdot 2(t-1) \binom{n-2}{k-2} = (t-1)^2 (3k-2) \binom{n-2}{k-2} + (t-1)(k-1) \cdot 2(t-1) \binom{n-2}{k-2} = (t-1)^2 (3k-2) \binom{n-2}{k-2} + (t-1)(k-1) \cdot 2(t-1) \binom{n-2}{k-2} = (t-1)^2 (3k-2) \binom{n-2}{k-2} + (t-1)(k-1) \cdot 2(t-1) \binom{n-2}{k-2} = (t-1)^2 (3k-2) \binom{n-2}{k-2} + (t-1)(k-1) \cdot 2(t-1) \binom{n-2}{k-2} = (t-1)^2 (3k-2) \binom{n-2}{k-2} + (t-1)(k-1) \cdot 2(t-1) \binom{n-2}{k-2} = (t-1)^2 (3k-2) \binom{n-2}{k-2} + (t-1)(k-1) \cdot 2(t-1) \binom{n-2}{k-2} = (t-1)^2 (3k-2) \binom{n-2}{k-2} + (t-1)(k-1) \cdot 2(t-1) \binom{n-2}{k-2} = (t-1)^2 (3k-2) \binom{n-2}{k-2} + (t-1)(k-1) \cdot 2(t-1) \binom{n-2}{k-2} = (t-1)^2 (3k-2) \binom{n-2}{k-2} + (t-1)(k-1) \cdot 2(t-1) \binom{n-2}{k-2} = (t-1)^2 (3k-2) \binom{n-2}{k-2} + (t-1)(k-1) \cdot 2(t-1) \binom{n-2}{k-2} = (t-1)^2 (3k-2) \binom{n-2}{k-2} + (t-1)(k-1) \cdot 2(t-1) \binom{n-2}{k-2} = (t-1)^2 (3k-2) \binom{n-2}{k-2} + (t-1)(k-1) \binom{n-2}{k-2} + (t-1)(k-1) \binom{n-2}{k-2} = (t-1)^2 \binom{n-2}{k-2} + (t-1)(k-1) \binom{n-2}{k-2} + (t-1)(k-1)$$

However, we know that the total number of edges exceeds

$$\begin{split} e(H) &> \binom{n}{k} - \binom{n-t+1}{k} \\ &= \left[1 - \left(1 - \frac{t-1}{n}\right) \cdots \left(1 - \frac{t-1}{n-k+1}\right)\right] \binom{n}{k} \\ &\ge \left[1 - \left(1 - \frac{t-1}{n}\right)^k\right] \binom{n}{k} \\ &\ge \left[k \cdot \frac{t-1}{n} - \binom{k}{2} \left(\frac{t-1}{n}\right)^2\right] \frac{n(n-1)}{k(k-1)} \binom{n-2}{k-2} \\ &\ge \left(\frac{(n-1)(t-1)}{k-1} - \frac{(t-1)^2}{2}\right) \binom{n-2}{k-2}, \end{split}$$

where we used that $(1-x)^k \leq 1 - kx + {k \choose 2}x^2$ when $0 \leq kx \leq 1$. Since $n > 3k^2t$, we also have n-1 > 3k(k-1)(t-1). Therefore,

$$e(H) > (t-1)^2 \left(3k - \frac{1}{2}\right) \binom{n-2}{k-2},$$

and so there is another edge in H disjoint from the previous t - 1 edges, again producing t disjoint edges in H.

Based on the same idea and technique, we can also obtain a multicolored version of the Erdős conjecture, which is an analogue of a theorem of Kleitman [54] for matching number greater than one. Note that Theorem 3.1.2 is the $\mathcal{F}_1 = \cdots = \mathcal{F}_t$ case of the following result.

Theorem 3.3.3. Let $\mathcal{F}_1, \ldots, \mathcal{F}_t$ be k-uniform families of subsets of [n], where $t < \frac{n}{3k^2}$, and every $|\mathcal{F}_i| > \binom{n}{k} - \binom{n-t+1}{k}$. Then there exist pairwise disjoint sets $F_1 \in \mathcal{F}_1, \ldots, F_t \in \mathcal{F}_t$.

Proof. For any vertex $v \in \mathcal{F}_i$, let H_v^j be the sub-hypergraph of \mathcal{F}_j induced by the vertex set $[n] \setminus \{v\}$. Then as in the previous proof,

$$e(H_v^j) \ge |\mathcal{F}_i| - \binom{n-1}{k-1} > \binom{n-1}{k} - \binom{(n-1) - (t-1) + 1}{k}.$$

By induction on t, for every i there exist t - 1 disjoint edges $\{e_j\}_{j \neq i}$ such that $e_j \in H_v^j$. So as before, if some \mathcal{F}_i has a vertex with degree $d(v) > k(t-1)\binom{n-2}{k-2}$, then there is an edge in \mathcal{F}_i which contains v and is disjoint from $\{e_j\}_{j \neq i}$. Hence we may assume the maximum degree in each hypergraph \mathcal{F}_i is at most $k(t-1)\binom{n-2}{k-2}$.

On the other hand, by induction on t we also know that for every i there exist t - 1 disjoint edges from the families $\{\mathcal{F}_j\}_{j\neq i}$, spanning (t-1)k vertices. If some \mathcal{F}_i has t-th largest degree at most $2(t-1)\binom{n-2}{k-2}$, then the sum of degrees of these (t-1)k vertices in \mathcal{F}_i is again at most

$$(t-1)^2(3k-2)\binom{n-2}{k-2} \le \binom{n}{k} - \binom{n-t+1}{k} < e(\mathcal{F}_i),$$

which guarantees the existence of an edge in \mathcal{F}_i disjoint from the previous t-1 edges from $\{\mathcal{F}_j\}_{j\neq i}$. So, we may assume that each \mathcal{F}_i contains at least t vertices with degree above $2(t-1)\binom{n-2}{k-2}$.

Now select distinct vertices v_i , such that for each $1 \leq i \leq t$, the degree of v_i in \mathcal{F}_i exceeds $2(t-1)\binom{n-2}{k-2}$. Consider all the subsets of $[n] \setminus \{v_1, \ldots, v_t\}$ which together with v_i form an edge of \mathcal{F}_i . Denote this (k-1)-uniform hypergraph by T^i . The same calculation as in Corollary 3.3.2 gives

$$e(T^i) > (t-1)\binom{n-t-1}{k-2}.$$

Applying Lemma 3.3.1 to $\{T^i\}$, we again find t disjoint edges from different families, as desired.

3.4 Concluding Remarks

In this chapter, we proved that for $t < \frac{n}{3k^2}$, every k-uniform hypergraph on n vertices with matching number less than t has at most $\binom{n}{k} - \binom{n-t+1}{k}$ edges. This verifies the conjecture

of Erdős in this range of t, and improves upon the previously best known range by a factor of k. As we discussed in the introduction, if the Erdős conjecture is true in general, then for $t < \frac{n}{k+1}$, the maximum number of edges cannot exceed $\binom{n}{k} - \binom{n-t+1}{k}$. It would be very interesting to tighten the range to $t < O(\frac{n}{k})$.

Pyber [79] proved the following product-type generalization of the Erdős-Ko-Rado theorem. Let \mathcal{F}_1 and \mathcal{F}_2 be families of k_1 - and k_2 -element subsets of [n]. If every pair of sets $F_1 \in \mathcal{F}_1$ and $F_2 \in \mathcal{F}_2$ intersects, then $|\mathcal{F}_1||\mathcal{F}_2| \leq {\binom{n-1}{k_1-1}} {\binom{n-1}{k_2-1}}$ for sufficiently large n. The special case when $k_1 = k_2$ and $\mathcal{F}_1 = \mathcal{F}_2$ corresponds to the Erdős-Ko-Rado theorem. Our Theorem 3.3.3 is a minimum-type result of similar flavor. Hence, it would be interesting to study the following multicolor analogue of Pyber's result.

Question 3.4.1. What is the maximum of $\prod_{i=1}^{t} |\mathcal{F}_i|$ among families $\mathcal{F}_1, \ldots, \mathcal{F}_t$ of subsets of [n], where each \mathcal{F}_i is k_i -uniform, and there are no t pairwise disjoint subsets $F_1 \in \mathcal{F}_1, \ldots, F_t \in \mathcal{F}_t$?

CHAPTER 4

Perfect Matching in hypergraphs

4.1 Introduction

Recall that in the previous chapter, we study the relation between the size of a hypergraph with its matching number. We defined a matching in a hypergraph H to be a set of disjoint edges of H. The number of edges in a matching is called *the size* of the matching. The size of the largest matching in a k-graph H is denoted by $\nu(H)$. A matching is *perfect* if its size equals |V|/k. Similarly, if the fractional matching number (see Section 2.3) $\nu^*(H) = n/k$, or equivalently, for all $v \in V$ we have $\sum_{e \ni v} w(e) = 1$, then we call w perfect.

As we mentioned earlier in Section 2.3, the determination of $\nu^*(H)$ is a linear programming problem. Its dual problem is to find a minimum fractional vertex cover $\tau^*(H) = \sum_{v \in V} w(v)$ over all functions $w : V \to [0, 1]$ such that for each $e \in E$ we have $\sum_{v \in e} w(v) \ge 1$. Let $\tau(H)$ be the minimum number of vertices in a vertex cover of H. Then, for every k-graph H, by the Duality Theorem,

$$\nu(H) \le \nu^*(H) = \tau^*(H) \le \tau(H). \tag{4.1.1}$$

Given a k-graph H and a set $S \in \binom{V}{d}$, $0 \le d \le k-1$, we denote by $\deg_H(S)$ the number of edges in H which contain S. Let $\delta_d := \delta_d(H)$ be the minimum d-degree of H, which is the minimum $\deg_H(S)$ over all $S \in \binom{V}{d}$. Note that $\delta_0(H) = |E(H)|$. In this chapter we study the relation between the minimum d-degree $\delta_d(H)$ and the matching numbers $\nu(H)$ and $\nu^*(H)$.

Definition 4.1.1. Let integers d, k, s, and n satisfy $0 \le d \le k - 1$, and $0 \le s \le n/k$. We denote by $m_d^s(k, n)$ the minimum m so that for an n-vertex k-graph H, $\delta_d(H) \ge m$ implies

that $\nu(H) \geq s$. Equivalently,

$$m_d^s(k,n) - 1 = \max\{\delta_d(H) : |V(H)| = n \text{ and } \nu(H) \le s - 1\}.$$

Furthermore, for a real number $0 \le s \le n/k$, define $f_d^s(k,n)$ as the minimum m so that $\delta_d(H) \ge m$ implies that $\nu^*(H) \ge s$. Equivalently,

$$f_d^s(k,n) - 1 = \max\{\delta_d(H) : |V(H)| = n \text{ and } \nu^*(H) < s\}.$$

Observe that trivially, for $\lceil s \rceil \leq n/k$,

$$f_d^s(k,n) \le m_d^{\lceil s \rceil}(k,n). \tag{4.1.2}$$

We are mostly interested in the case s = n/k (i.e. when matchings are perfect) in which we suppress the superscript in the notation $m_d^{n/k}(k,n)$ and $f_d^{n/k}(k,n)$. Thus, writing $m_d(k,n)$, we implicitly require that n is divisible by k.

Problems of this type have a long history going back to Dirac [27] who in 1952 proved that minimum degree n/2 implies the existence of a Hamiltonian cycle in graphs. Therefore, for $d \ge 1$, we refer to the extremal parameters $m_d(k, n)$ and $f_d(k, n)$ as to Dirac-type thresholds. When k = 2, an easy argument shows that $m_1(2, n) = n/2$. For $k \ge 3$, an exact formula for $m_{k-1}(k, n)$ was obtained in [85]. For a fixed $k \ge 3$ and $n \to \infty$ it yields the asymptotics $m_{k-1}(k, n) = \frac{n}{2} + O(1)$. As far as perfect fractional matchings are concerned, it was proved in [83] that $f_{k-1}(k, n) = \lceil n/k \rceil$ for $k \ge 2$, which is a lot less than $m_{k-1}(k, n)$ when $k \ge 3$. For more results on Dirac-type thresholds for matchings and Hamilton cycles see [82].

In this chapter, we focus on the asymptotic behavior of $m_d(k, n)$ and $f_d(k, n)$ for general, but fixed k and d, when $n \to \infty$. For a lower bound on $m_d(k, n)$ consider first a k-graph $H_0 = H_0(k, n)$ (constructed in [85]) with vertex set split almost evenly, that is, $V(H_0) = A \cup B$, $||A| - |B|| \le 2$, and with the edge set consisting of all k-element subsets of $V(H_0)$ intersecting A in an odd number of vertices. We choose the size of A so that |A| and $\frac{n}{k}$ have different parity. Clearly, there is no perfect matching in H_0 and for every $0 \le d \le k-1$ we have $\delta_d(H_0) \sim \frac{1}{2} {n-d \choose k-d}$.

Another lower bound on $m_d(k, n)$ is given by the following well-known construction, which was mentioned earlier in Section 3.1, as one of the two conjectured extremal examples for the Erdős conjecture. For integers n, k, and s, let $H_1(s)$ be a k-graph on n vertices consisting of all k-element subsets intersecting a given set of size s - 1, that is $H_1(s) = K_n^{(k)} - K_{n-s+1}^{(k)}$. Observe that $\nu(H_1(s)) = s - 1$, while

$$\delta_d(H_1(n/k)) = \binom{n-d}{k-d} - \binom{n-d-n/k+1}{k-d} \sim \left\{ 1 - \left(\frac{k-1}{k}\right)^{k-d} \right\} \binom{n-d}{k-d}.$$

Assume that n is divisible by k. Putting $s = \frac{n}{k}$ and using the k-graphs H_0 and $H_1(n/k)$, we obtain a lower bound

$$m_d(k,n) \ge \max\left\{\delta_d(H_0), \delta_d(H_1(\frac{n}{k}))\right\} + 1 \sim \max\left\{\frac{1}{2}, 1 - \left(\frac{k-1}{k}\right)^{k-d}\right\} \binom{n-d}{k-d}.$$
 (4.1.3)

On the other hand, $H_1(\lceil n/k \rceil)$ alone yields a lower bound also on $f_d(k, n)$. Indeed, for a real s > 0 we have

$$\nu^*(H_1(\lceil s \rceil)) = \tau^*(H_1(\lceil s \rceil)) \le \tau(H_1(\lceil s \rceil)) = \lceil s \rceil - 1 < s,$$

and so

$$f_d(k,n) \ge \delta_d(H_1(\lceil \frac{n}{k} \rceil)) + 1 \sim \left\{ 1 - \left(\frac{k-1}{k}\right)^{k-d} \right\} \binom{n-d}{k-d}.$$
(4.1.4)

It is easy to check that for $d \ge k/2$, the maximum in the coefficient in (4.1.3) equals $\frac{1}{2}$. Pikhurko [78] proved, complementing the case d = k - 1, that indeed we have $m_d(k,n) \sim \frac{1}{2} \binom{n-d}{k-d}$ also for $k/2 \le d \le k - 2$, $k \ge 4$.

For d < k/2 the problem seems to be harder and we discuss below the cases $d \ge 1$ and d = 0 separately. The first result for the range $1 \le d < k/2$, $k \ge 3$, was obtained already in 1981 by Daykin and Häggkvist in [25] who proved that $m_1(k,n) \le \left(\frac{k-1}{k} + o(1)\right) \binom{n-1}{k-1}$. This was generalized to $m_d(k,n) \le \left(\frac{k-d}{k} + o(1)\right) \binom{n-d}{k-d}$ for all $1 \le d < k/2$ in [43], and, using the ideas from [43], slightly improved in [71] to $m_d(k,n) \le \left\{\frac{k-d}{k} - \frac{1}{k^{k-d}} + o(1)\right\} \binom{n-d}{k-d}$. For k = 4, d = 1 the latter coefficient is $\frac{47}{64}$. In [71], the constant was further lowered to $\frac{42}{64}$, but there is still gap between this upper bound and the lower bound of $\frac{37}{64}$.

It has been conjectured in [57] and again in [43] that the lower bound (4.1.3) is achieved at least asymptotically.

Conjecture 4.1.2. For all $1 \le d \le k-1$,

$$m_d(k,n) \sim \max\left\{\frac{1}{2}, 1 - \left(\frac{k-1}{k}\right)^{k-d}\right\} \binom{n-d}{k-d}.$$

Hàn, Person, and Schacht in [43] proved Conjecture 4.1.2 in the case d = 1, k = 3 by showing that $m_1(3, n)$ is asymptotically equal to $\frac{5}{9} \binom{n-1}{2}$. Kühn, Osthus, and Treglown [58] and, independently, Khan [52], proved the exact result $m_1(3, n) = \delta_1(H_1(n/3)) + 1$. Recently Khan [53] announced that he verified the exact result $m_1(4, n) = \delta_1(H_1(n/4)) + 1$, while the asymptotic version, $m_1(4, n) \sim \frac{37}{64} \binom{n-1}{3}$ follows also from a more general result by Lo and Markström [64].

These exact results, together with (4.1.2) and (4.1.4), yield that $f_1(3,n) = m_1(3,n)$ and $f_1(4,n) = m_1(4,n)$. Remembering that, on the other hand, $f_{k-1}(k,n)$ is much smaller than $m_{k-1}(k,n)$, one can raise the question about a general relation between $m_d(k,n)$ and its fractional counterpart $f_d(k,n)$. In this chapter we answer this question by showing that $m_d(k,n)$ and $f_d(k,n)$ are asymptotically equal whenever $f_d(k,n) \sim c^* \binom{n-d}{k-d}$ for some constant $c^* > \frac{1}{2}$, and otherwise $m_d(k,n) \sim \frac{1}{2} \binom{n-d}{k-d}$.

Theorem 4.1.3. For every $1 \le d \le k-1$ if there exists $c^* > 0$ such that $f_d(k, n) \sim c^* \binom{n-d}{k-d}$ then

$$m_d(k,n) \sim \max\left\{c^*, \frac{1}{2}\right\} \binom{n-d}{k-d}.$$
 (4.1.5)

This result reduces the task of asymptotically calculating $m_d(k, n)$ to a presumably simpler task of calculating $f_d(k, n)$. It seems that, similarly to the integral case, the lower bound in (4.1.4) determines asymptotically the actual value of the parameter $f_d(k, n)$.

Conjecture 4.1.4. For all $1 \le d \le k-1$,

$$f_d(k,n) \sim \left\{ 1 - \left(\frac{k-1}{k}\right)^{k-d} \right\} \binom{n-d}{k-d}.$$

Our next result confirms Conjecture 4.1.4 asymptotically for all k and d such that $1 \leq k - d \leq 4$. Note that the above mentioned result from [83] shows that Conjecture 4.1.4 is true for d = k - 1 exactly, that is, $f_{k-1}(k,n) = \delta_{k-1}\left(H_1\left(\lceil \frac{n}{k} \rceil\right)\right) + 1$. We include this case into the statement of Theorem 4.1.5 for completeness.

Theorem 4.1.5. For every $k \ge 3$ and $k - 4 \le d \le k - 1$, we have

$$f_d(k,n) \sim \left\{ 1 - \left(\frac{k-1}{k}\right)^{k-d} \right\} \binom{n-d}{k-d}.$$

Theorems 4.1.5 and 4.1.3 together imply immediately the validity of Conjecture 4.1.2 in a couple of new instances (as discussed earlier, the first of them has been recently also proved in [53] and [64]).

Corollary 4.1.6. We have

$$m_1(4,n) \sim \frac{37}{64} \binom{n-1}{3}, \quad m_2(5,n) \sim \frac{1}{2} \binom{n-2}{3}, \quad m_1(5,n) \sim \frac{369}{625} \binom{n-1}{4}$$

 $m_2(6,n) \sim \frac{671}{1296} \binom{n-2}{4}, \quad m_3(7,n) \sim \frac{1}{2} \binom{n-3}{4}.$

We prove Theorem 4.1.5 utilizing the following connection between the parameters $f_d^s(k,n)$ and $f_0^s(k-d,n-d)$.

Proposition 4.1.7. For all $k \ge 3$, $1 \le d \le k - 1$, and $n \ge k$,

$$f_d(k,n) \le f_0^{n/k}(k-d,n-d).$$

In view of Proposition 4.1.7, in order to prove Theorem 4.1.5 we need to estimate $f_0^s(k-d, n-d)$ with $s = \frac{n}{k}$. This is trivial for d = k - 1 and so, from now on, we will be assuming that $d \le k - 2$. The integral version of this problem has almost as long history as the Dirac-type problem $(d \ge 1)$.

Erdős and Gallai [29] determined $m_0^s(k, n)$ for graphs (k = 2). And as mentioned earlier in Chapter 2 and 3, Erdős [28] conjectured the following hypergraph generalization of their result. **Conjecture 4.1.8.** For all $k \ge 2$ and $1 \le s \le \frac{n}{k}$:

$$m_0^s(k,n) = \max\left\{\binom{ks-1}{k}, \binom{n}{k} - \binom{n-s+1}{k}\right\} + 1.$$

The lower bound comes from considering again the extremal k-graph $H_1(s)$ along with the k-uniform clique $K_{ks-1}^{(k)}$ (complemented by n - ks + 1 isolated vertices) which, clearly, has no matching of size s. For more on Erdős' conjecture we refer the reader to the survey paper [34] and a recent paper [36], where the conjecture is proved for k = 3 and $n \ge 4s$. In its full generality, the conjecture is still wide open.

We now formulate the fractional version of Erdős' Conjecture. For future references, we switch from k and n to l and m. Again, the lower bound is yielded by $H_1(\lceil s \rceil)$ and the complete l-graph on $\lceil ls \rceil - 1$ vertices, $K_{\lceil ls \rceil - 1}^{(l)}$.

Conjecture 4.1.9. For all integers $l \ge 2$ and an integer s such that $0 \le s \le m/l$, we have

$$f_0^s(l,m) = \max\left\{ \binom{\lceil ls \rceil - 1}{l}, \binom{m}{l} - \binom{m - \lceil s \rceil + 1}{l} \right\} + 1.$$

Note that Conjecture 4.1.9 implies that the bound is also asymptotically true for noninteger values of s, when m is large. In [63], there is an example showing that the stronger, precise version of the conjecture does not hold for fractional s.

As a consequence of the Erdős-Gallai theorem from [29], Conjecture 4.1.9 is asymptotically true for l = 2 and m goes to infinity. In the next section we establish a result which confirms Conjecture 4.1.9 asymptotically in the two smallest new instances, but limited to the range $0 \le s \le \frac{m}{l+1}$. In this range the case l = 3 follows also from the above mentioned result in [36]. It is easy to check that for $s \le \frac{m}{l+1} + O(1)$, the maximum in Conjecture 4.1.9 is achieved by the second term.

Theorem 4.1.10. For $l \in \{3, 4\}$, for all $d \ge 1$, and $s = \frac{m+d}{l+d}$,

$$f_0^s(l,m) \sim \left\{ 1 - \left(1 - \frac{1}{l+d}\right)^l \right\} \binom{m}{l}$$

where the asymptotics holds for $m \to \infty$ with d fixed.

Theorem 4.1.10 together with Proposition 4.1.7 implies Theorem 4.1.5, which, in turn, together with Theorem 4.1.3 yields Corollary 4.1.6. To prove Conjecture 4.1.2 in full generality, one would need to prove Theorem 4.1.10 for all l.

The rest of this chapter is organized as follows. In the next section, we prove Theorem 4.1.10 using as a main tool a probabilistic inequality of Samuels. A proof of Proposition 4.1.7, and consequently of Theorem 4.1.5, appears in Section 4.3. Section 4.4 contains a proof of Theorem 4.1.3. Finally, in Section 4.5, we discuss an application of the fractional version of the Erdös problem in distributed storage allocation. The last section contains concluding remarks and open problems.

4.2 Fractional matchings and probability of small deviations

In this section we prove Theorem 4.1.10 using a probabilistic approach from [?] based on a special case of an old probabilistic conjecture of Samuels [86], which we briefly mentioned in Section 3.4. In fact, we prove a little bit more – see Corollary 4.2.5 and Remark 4.2.6 below.

For *l* reals μ_1, \ldots, μ_l satisfying $0 \le \mu_1 \le \mu_2 \le \cdots \le \mu_l$ and $\sum_{i=1}^l \mu_i < 1$, let

$$P(\mu_1, \mu_2, \ldots, \mu_l) = \inf \mathbb{P}(X_1 + \ldots + X_l < 1),$$

where the infimum is taken over all possible collections of l independent nonnegative random variables X_1, \ldots, X_l , with expectations μ_1, \ldots, μ_l , respectively. Define

$$Q_t(\mu_1, \dots, \mu_l) = \prod_{i=t+1}^l \left(1 - \frac{\mu_i}{1 - \sum_{j=1}^t \mu_j} \right)$$

for each $0 \le t < l$.

Note that $Q_t(\mu_1, \ldots, \mu_l)$ is exactly $\mathbb{P}(X_1 + \ldots + X_l < 1)$ when X_i is identically μ_i for all $i \leq t$, while X_i attains the values 0 and $1 - \sum_{i \leq t} \mu_i$ (with its expectation being μ_i) for all $i \geq t+1$.

The following conjecture was raised by Samuels in [86].

Conjecture 4.2.1 ([86]). For all admissible values of μ_1, \ldots, μ_l ,

$$P(\mu_1, \mu_2, \dots, \mu_l) = \min_{t=0,\dots,l-1} Q_t(\mu_1, \mu_2, \dots, \mu_l).$$

Note that for l = 1 this is Markov's inequality. Samuels proved his conjecture for $l \le 4$ in [86, 87].

Lemma 4.2.2 ([86, 87]). The assertion of Conjecture 4.2.1 holds for all $l \leq 4$.

We next show that for $\mu_1 = \mu_2 = \cdots = \mu_l = x$, where $0 < x \leq \frac{1}{l+1}$, the minimum in Conjecture 4.2.1 is attained by $Q_0(\mu_1, \ldots, \mu_l)$.

Proposition 4.2.3. For every integer $l \ge 2$ and every real number x satisfying $0 < x \le \frac{1}{l+1}$, if $\mu_1 = \mu_2 = \ldots = \mu_l = x$ then

$$\min_{t=0,\dots,l-1} Q_t(\mu_1,\mu_2,\dots,\mu_l) = Q_0(\mu_1,\mu_2,\dots,\mu_l) = (1-x)^l.$$

Proof. By definition

$$Q_t(\mu_1,\mu_2,\ldots,\mu_l) = \left(1 - \frac{x}{1 - tx}\right)^{l-t} = \left(\frac{1 - (t+1)x}{1 - tx}\right)^{l-t}.$$

We thus have to prove that for $0 < x \leq \frac{1}{l+1}$ and $1 \leq t \leq l-1$,

$$(1-x)^{l} \le \left(\frac{1-(t+1)x}{1-tx}\right)^{l-t}$$

or equivalently that

$$\left(\frac{1}{1-x}\right)^l \ge \left(\frac{1-tx}{1-(t+1)x}\right)^{l-t}.$$

By the geometric-arithmetic means inequality applied to a set of l numbers, t of which are equal to 1 and the remaining l - t equal to the quantity $\frac{1-tx}{1-(t+1)x}$, we conclude that

$$\left(\frac{1-tx}{1-(t+1)x}\right)^{l-t} \cdot 1^t \le \left[\frac{1}{l} \cdot \left(\frac{(1-tx)(l-t)}{1-(t+1)x} + t\right)\right]^l.$$

Thus, it suffices to show that

$$\frac{(1-tx)(l-t)}{1-(t+1)x} + t \le \frac{l}{1-x}.$$

This is equivalent to

$$(1-x)[(1-tx)(l-t)+t-t(t+1)x] \le l[1-(t+1)x],$$

which is equivalent to

$$(1-x)[l-t(l+1)x] \le l - l(t+1)x,$$

or

$$l - t(l+1)x - lx + t(l+1)x^{2} \le l - l(t+1)x.$$

After dividing by x, we see that this is equivalent to $x \leq \frac{1}{l+1}$, which holds by assumption, completing the proof.

Note that when s = xm and $x \leq \frac{1}{l+1}$, the maximum in Conjecture 4.1.9 is achieved by the second term. We now prove the following, in most part conditional, result, which shows how to deduce Conjecture 4.1.9 in this range from Conjecture 4.2.1.

Theorem 4.2.4. For any $l \ge 3$ and $0 < x \le \frac{1}{l+1}$, if Conjecture 4.2.1 holds for l and $\mu_1 = \mu_2 = \ldots = \mu_l = x$ then

$$f_0^{xm}(l,m) \sim \left\{1 - (1-x)^l\right\} \binom{m}{l}.$$

Combining Theorem 4.2.4 with Lemma 4.2.2, we obtain the following corollary which implies Theorem 4.1.10. (For d = 1, observe that $f_0^s(l,m) \sim f_0^s(l,m+1)$.)

Corollary 4.2.5. For l = 3, $x \le 1/4$ and for l = 4, $x \le 1/5$, the maximum number of edges in an *l*-uniform hypergraph *H* on *m* vertices with fractional matching number less than xm is

$$f_0^{xm}(l,m) \sim \left\{1 - (1-x)^l\right\} \binom{m}{l}.$$

Proof of Theorem 4.2.4: Let H be an l-uniform hypergraph on a vertex set V, |V| = m, and suppose that $\nu^*(H) < xm$. By duality, we also have $\tau^*(H) < xm$, and hence there exists a weight function $w: V \to [0, 1]$ such that $\sum_{v \in V} w(v) < xm$ and, for every edge e of H, $\sum_{v \in e} w(v) \ge 1$. By increasing the weights w(v) if needed, we may assume that

$$\sum_{v \in V}^{m} w(v) = xm$$

Let v_1, \ldots, v_l be a sequence of random vertices of H, chosen independently and uniformly at random from V. For each $i = 1, \ldots, l$ we define a random variable $X_i = w(v_i)$. Note that X_1, X_2, \ldots, X_l are independent, identically distributed random variables, where every X_i attains each of the *m* values w(v) with probability 1/m. (The values of *w* for different vertices can be equal, but this is of no importance for us.)

By definition, the expectation μ_i of each X_i is

$$\mu_i = \sum_{v \in V} \frac{1}{m} \cdot w(v) = \frac{xm}{m} = x.$$

Now we can estimate the number of edges of H as follows. Since for each edge of H we have $\sum_{v \in e} w(v) \ge 1$, the number N of all l-element subsets S of V with $\sum_{v \in S} w(v) < 1$ is a lower bound on the number of non-edges of H. Let N_1 and N_2 be the numbers of all l-element sequences of vertices of V and all l-element sequences of distinct vertices of V, respectively, with the sums of weights strictly smaller than 1. Note that $N_1 - N_2$ is at most the number of l-element sequences in which at least one vertex appears twice, thus it is bounded by $\binom{l}{2}m^{l-1} = O(m^{l-1})$. As the number of all l-element subsets of V is $\binom{m}{l} = (1 + o(1))m^l/l!$ and $N = N_2/l!$, we have

$$\mathbb{P}\left(\sum_{i=1}^{l} w(v_i) < 1\right) = \frac{N_1}{m^l} \le \frac{N_2 + O(m^{l-1})}{\binom{m}{l}l!} = (1 + o(1))\frac{N}{\binom{m}{l}}$$

If Conjecture 4.2.1 holds for a given l then, by Lemma 4.2.2 and Proposition 4.2.3,

$$\mathbb{P}\left(\sum_{i=1}^{l} w(v_i) < 1\right) = \mathbb{P}\left(\sum_{i=1}^{l} X_i < 1\right) \ge (1-x)^l,$$

and, consequently,

$$N \ge (1+o(1))(1-x)^l \binom{m}{l}.$$

It follows that the number of edges of H is at most $(1 + o(1)) \{1 - (1 - x)^l\} \binom{m}{l}$, as needed.

Remark 4.2.6. Note that the above proof works as long as the conclusion of Proposition 4.2.3 holds. One can check using Mathematica that Proposition 4.2.3 holds for l = 3 and all $0 < x \le 0.277$, as well as for l = 4 and all $0 < x \le 0.217$. Therefore, Corollary 4.2.5 extends to these broader ranges of x. For bigger values of x, e.g., for x = 0.3 when l = 3, this is not the case anymore, and the above method does not suffice to determine the asymptotic

behavior of $f_0^{xm}(l,m)$. In fact, using Samuels conjecture in the higher range of x, one gets a bound on $f_0^{xm}(l,m)$ which is larger than that in Conjecture 4.1.9. However, in view of Proposition 4.1.7, for our main application the case $x \leq \frac{1}{l+1}$ is just what we need.

4.3 Thresholds for perfect fractional matchings

In this section we present a proof of Proposition 4.1.7 and then deduce quickly Theorem 4.1.5.

Proof of Proposition 4.1.7: The outline of the proof goes as follows. We will assume that there is no fractional perfect matching in a k-graph H on n vertices and then show that the neighborhood graph H(L) in H of a particular set L of size d satisfies $\nu^*(H(L)) < n/k$. This will imply that $\delta_d(H) \leq |H(L)| < f_0^{n/k}(k-d,n-d)$. In contrapositive, we will prove that if $\delta_d(H) \geq f_0^{n/k}(k-d,n-d)$ then H has a fractional perfect matching, from which it follows, by definition, that $f_d(k,n) \leq f_0^{n/k}(k-d,n-d)$.

Let an *n*-vertex *k*-graph *H* satisfy $\nu^*(H) < n/k$, that is, have no fractional perfect matching. As $\tau^*(H) = \nu^*(H)$, there is a function $w : V \to [0, 1]$ such that $\sum_{v \in V} w(v) < n/k$ and, for every $e \in H$, we have $\sum_{v \in e} w(v) \ge 1$. We can replace *H* with the *k*-graph whose edge set consists of every *k*-tuple of vertices on which *w* totals to at least one.

Formally, for every weight function $w:V\rightarrow [0,1]$ define

$$H_w := \left\{ e \in \binom{V}{k} : \sum_{v \in e} w(v) \ge 1 \right\}.$$

For a given weight function w, suppose L is a set of d vertices with the smallest weights. Without loss of generality, we may assume that the d lowest values of w(x) are all equal to each other, since otherwise we could replace them by their average. (Obviously, this modification would not change $\sum_{v \in V} w(v)$ nor the set L.) Note that the minimum d-degree $\delta_d(H_w) = \min_{S \subset \binom{V}{d}} \deg_H(S)$ is achieved when S = L. Let H(L) be the neighborhood of Lin H_w , that is a (k - d)-graph on the vertex set $V \setminus L$ and with the edge set

$$\left\{S \in \binom{V-L}{k-d} : S \cup L \in E(H_w)\right\}.$$

Then $|H(L)| = \delta_d(H_w)$ and it remains to prove that $\tau^*(H(L)) < n/k$.

Let $w_0 = \min_{v \in V} w(v)$ and observe that $w_0 < 1/k$. If $w_0 > 0$, apply to the weight function w the following linear map

$$w' = \frac{w - w_0}{1 - kw_0}.$$

Then, still $\sum_{v \in V} w'(v) < n/k$ and $H_w = H_{w'}$. Moreover, for every $v \in L$, we have w'(v) = 0. It follows that the function w' restricted to the set $V \setminus L$ is a fractional vertex cover of H(L) and so $\nu^*(H(L)) = \tau^*(H(L)) < n/k$, which completes the proof of Proposition 4.1.7. \Box **Proof of Theorem 4.1.5**: As explained earlier, $f_0^{n/k}(k-d, n-d) = n/k$ holds trivially for d = k - 1 and together with Proposition 4.1.7 implies the theorem in this case. For d = k - 2, we apply Proposition 4.1.7 together with the case l = 2 of the fractional Erdős Conjecture 4.1.9 (as mentioned earlier, it follows asymptotically from [29]). For d = k - 3 and d = k - 4, we use Proposition 4.1.7 and Corollary 4.1.10 proved in the previous section.

Remark 4.3.1. Consider a restricted version of Samuels' problem to minimize $\mathbb{P}(X_1 + \cdots + X_l < 1)$ under the additional assumption that all random variables are *i*-dentically distributed. Our proofs indicate that under this regime, for a given $l \ge 5$ and $\mu_1 = \cdots = m_l = x \le \frac{1}{l+1}$, if

$$\mathbb{P}(X_1 + \dots + X_l < 1) \ge (1 + o(1))(1 - x)^l$$

then Theorem 4.1.5 would hold for all $k \ge l+1$ and d = k-l.

4.4 Constructing integer matchings from fractional ones

In this section, we will prove Theorem 4.1.3. An indispensable tool in our proof is the Strong Absorbing Lemma 4.4.1 from [43] (see Lemma 10 therein). This lemma provides a sufficient condition on degrees and co-degrees of a hypergraph ensuring the existence of a small and powerful matching which, by "absorbing" vertices, creates a perfect matching from any nearly perfect matching.

Lemma 4.4.1. For all $\gamma > 0$ and integers k > d > 0 there is an n_0 such that for all $n > n_0$

the following holds: suppose that H is a k-graph on n vertices with $\delta_d(H) \ge (1/2 + 2\gamma) {n-d \choose k-d}$, then there exists a matching $M := M_{abs}$ in H such that

- (i) $|M| < \gamma^k n/k$, and
- (ii) for every set $W \subset V \setminus V(M)$ of size at most $|W| \leq \gamma^{2k}n$ and divisible by k there exists a matching in H covering exactly the vertices of $V(M) \cup W$.

Equipped with this lemma we can practically reduce our task to finding an almost perfect matching in a suitable subhypergraph of H. Here is an outline of our proof of Theorem 4.1.3. Assume that there exists a constant $0 < c^* < 1$ such that $f_d(k, n) \sim c^* \binom{n-d}{k-d}$. For any $\alpha > 0$ consider a k-graph H on n vertices, where n is sufficiently large, with

$$\delta_d(H) \ge (c+\alpha) \binom{n-d}{k-d}$$

where $c = \max\{\frac{1}{2}, c^*\}$. Our goal is to show that H contains a perfect matching.

Set $\gamma = \alpha/2$ and $\varepsilon = \gamma^{2k}$. The proof consists of three steps.

1. Find an absorbing matching M_{abs} satisfying properties (i) and (ii) of Lemma 4.4.1. Set $H' = H \setminus V(M_{abs})$ and note that when n is sufficiently large,

$$\delta_d(H') \ge \delta_d(H) - \left(\binom{n-d}{k-d} - \binom{n-d-\varepsilon n}{k-d}\right) \ge (c+\alpha/2)\binom{n-d}{k-d} = (c+\gamma)\binom{n-d}{k-d}.$$

- 2. Find a matching M_{alm} in H' such that $|V(M_{alm})| \ge (1 \varepsilon)|V(H')|$, and thus, $|V(M_{alm} \cup M_{abs})| \ge (1 - \varepsilon)n.$
- 3. Extend $M_{alm} \cup M_{abs}$ to a perfect matching of H by using the absorbing property (ii) of M_{abs} with respect to $W = V(H') \setminus V(M_{alm})$.

Now come the details of the proof. The Strong Absorbing Lemma provides an absorbing matching M_{abs} , so Steps 1 and 3 are clear. Hence to complete the proof of Theorem 4.1.3 it remains to explain Step 2. One possible approach to find an almost perfect matching in H' is via the weak hypergraph regularity lemma. Our proof, however, is based on Theorem 1.1 in [35]. Recall that the 2-degree of a pair of vertices in a hypergraph is the number of

edges containing this pair. An immediate corollary of that theorem asserts the existence of an almost perfect matching in any nearly regular k-graph in which all 2-degrees are much smaller than the vertex degrees. (See Remark after Theorem 1.1 in [35] or Chapter 4.7 of [6]). Here we formulate this corollary as the following lemma in which $\Delta_2(H)$ denotes the maximum 2-degree in H.

Lemma 4.4.2. For every integer $k \ge 2$ and a real $\varepsilon > 0$, there exists $\tau = \tau(k, \varepsilon)$, $d_0 = d_0(k, \varepsilon)$ such that for every $n \ge D \ge d_0$ the following holds.

Every k-uniform hypergraph on a set V of n vertices which satisfies the following conditions:

- 1. $(1 \tau)D < \deg_H(v) < (1 + \tau)D$ for all $v \in V$, and
- 2. $\Delta_2(H) < \tau D$

contains a matching M_{alm} covering all but at most εn vertices.

Hence, Step 2 above reduces to finding a spanning subhypergraph H'' of H' satisfying the assumptions of Lemma 4.4.2 with $\varepsilon = \gamma^{2k}$ and other parameters τ, D, a to be suitably chosen. Indeed, the following claim is all we need to complete the proof of Theorem 4.1.3. For convenience, we set n := |V(H')|. Recall that $c = \max\{\frac{1}{2}, c^*\}$ where c^* comes from the threshold which guarantees the existence of fractional perfect matchings.

Claim 4.4.3. For sufficiently large n, any k-graph H' on n vertices satisfying $\delta_d(H') \ge (c+\gamma) \binom{n-d}{k-d}$ contains a spanning subhypergraph H'', such that for all $v \in V(H'')$ we have $\deg_{H''}(v) \sim n^{0.2}$ while $\Delta_2(H'') \le n^{0.1}$.

Consequently for every $k \ge 2$, $\varepsilon > 0$, the subhypergraph H'' satisfies the assumptions of Lemma 4.4.2 with $D = n^{0.2}$, and any $\tau > 0$. We obtained the following result as an immediate corollary, which asserts the validity of Step 2 and completes our proof of Theorem 4.1.3.

Corollary 4.4.4. *H'* contains an almost perfect matching covering at least $(1 - \varepsilon)|V(H')|$ vertices.

In the proof of Claim 4.4.3, the following well-known concentration results (see, for example [6], Appendix A, and Theorem 2.8, inequality (2.9) and (2.11) in [49]) will be used several times. We denote by Bi(n, p) a binomial random variable with parameters n and p.

Lemma 4.4.5. (Chernoff Inequality for small deviation) If $X = \sum_{i=1}^{n} X_i$, each random variable X_i has Bernoulli distribution with expectation p_i , and $\alpha \leq 3/2$, then

$$\mathbb{P}(|X - \mathbb{E}X| \ge \alpha \mathbb{E}X) \le 2e^{-\frac{\alpha^2}{3}\mathbb{E}X}$$
(4.4.1)

In particular, when $X \sim Bi(n,p)$ and $\lambda < \frac{3}{2}np$, then

$$\mathbb{P}(|X - np| \ge \lambda) \le e^{-\Omega(\lambda^2/(np))} \tag{4.4.2}$$

Lemma 4.4.6. (Chernoff Inequality for large deviation) If $X = \sum_{i=1}^{n} X_i$, each random variable X_i has Bernoulli distribution with expectation p_i , and $x \ge 7 \mathbb{E}X$, then

$$\mathbb{P}(X \ge x) \le e^{-x} \tag{4.4.3}$$

Proof of Claim 4.4.3: The desired subhypergraph H'' is obtained via two rounds of randomization. In the first round, we find edge-disjoint induced subhypergraphs with large minimum degrees which guarantees the existence of perfect fractional matchings. In the second round, we construct H'' from these fractional matchings.

As a preparation toward the first round, R is obtained by choosing every vertex randomly and independently with probability $p = |V'|^{-0.9} = n^{-0.9}$. Then |R| is a binomial random variable with expectation $n^{0.1}$. By inequality (4.4.2), $|R| \sim n^{0.1}$ with probability $1 - e^{-\Omega(n^{0.1})}$.

Fix a subset $D \subseteq V'$ of size d and let DEG_D be the number of edges $f \in H'$ such that $D \subset f$ and $f \setminus D \subseteq R$, which is the number of edges e in the link graph H[D] with all of its vertices in the random set R. Therefore $DEG_D = \sum_{e \in H[D]} X_e$, where $X_e = 1$ if e is in R and 0 otherwise. We have

$$\mathbb{E}(\text{DEG}_D) = \deg_{H'}(D) \times (n^{-0.9})^{k-d} \ge (c + \alpha/2) \binom{n-d}{k-d} n^{-0.9(k-d)}$$
$$\ge (c + \alpha/3) \binom{|R| - d}{k-d} = \Omega(n^{0.1(k-d)})$$

For two distinct intersecting edges e_i, e_j with $|e_i \cap e_j| = l$ for $1 \le l \le k-d-1$, the probability that both of them are in R is

$$\mathbb{P}(X_{e_i} = X_{e_j} = 1) = p^{2(k-d)-l}$$

For fixed l, there are at most $\binom{n-d}{k-d}$ choices for e_i in the link graph H[D], $\binom{k-d}{l}$ ways to choose the intersection $L = e_i \cap e_j$ of size l, and $\binom{(n-d)-(k-d)}{k-d-l}$ options for $e_j \setminus L$. Therefore,

$$\Delta = \sum_{e_i \cap e_j \neq \emptyset} \mathbb{P}(X_{e_i} = X_{e_j} = 1) \le \sum_{l=1}^{k-d-1} p^{2(k-d)-l} \binom{n-d}{k-d} \binom{k-d}{l} \binom{n-k}{k-d-l} \le \sum_{l=1}^{k-d-l} p^{2(k-d)-l}O(n^{2(k-d)-l}) = O(n^{0.1(2(k-d)-1)})$$

By Janson's inequality (see Theorem 8.7.2 in [6]),

$$\mathbb{P}(\mathrm{DEG}_D \le (1 - \alpha/12)\mathbb{E}(\mathrm{DEG}_D)) \le e^{-\Omega((\mathbb{E}X)^2/\Delta)} \sim e^{-\Omega(n^{0.1})}$$

Therefore by the union bound, with probability $1 - n^d e^{-\Omega(n^{0.1})}$, for all subsets $D \subseteq V'$ of size d, we have

$$\operatorname{DEG}_D > (1 - \alpha/12) \mathbb{E}(\operatorname{DEG}_D)) \ge (c + \alpha/4) \binom{|R| - d}{k - d}$$

Take $n^{1.1}$ independent copies of R and denote them by R^i , $1 \le i \le n^{1.1}$, and the corresponding random variables by $\text{DEG}_D^{(i)}$, where $D \subseteq V'$, |D| = d, and $i = 1, \ldots, n^{1.1}$. Since $|R_i| \sim n^{0.1}$ with probability $1 - e^{-\Omega(n^{0.1})}$ for each i, the union bound ensures that $|R_i| \sim n^{0.1}$ for every $i = 1, \cdots, n^{1.1}$ with probability 1 - o(1). Now for a subset of vertices $S \subseteq V'$, define the random variable

$$Y_S = |\{i : S \subseteq R^i\}|.$$

Note that the random variables Y_S have binomial distributions $Bi(n^{1.1}, n^{-0.9|S|})$ with expectations $n^{1.1-0.9|S|}$. In particular, for every vertex $v \in V'$, $Y_{\{v\}} \sim Bi(n^{1.1}, n^{-0.9})$ and $\mathbb{E}Y_{\{v\}} = n^{0.2}$. Hence, by inequality (4.4.2), taking $\lambda = n^{0.15}$,

$$\mathbb{P}(|Y_{\{v\}} - n^{0.2}| > n^{0.15}) \le e^{-\Omega((n^{0.15})^2/n^{0.2})} = e^{-\Omega(n^{0.1})}$$

Therefore a.a.s $|Y_{\{v\}} - n^{0.2}| \le n^{0.15}$ for every vertex $v \in V'$.

Further, let

$$Z_{2} = \left| \left\{ \{u, v\} \in \binom{V'}{2} : Y_{\{u, v\}} \ge 3 \right\} \right|.$$

Then

$$\mathbb{E}Z_2 < n^2 (n^{1.1})^3 (n^{-0.9})^6 = n^{-0.1}.$$

Therefore by Markov's inequality,

$$\mathbb{P}(Z_2 = 0) = 1 - \mathbb{P}(Z_2 \ge 1) \ge 1 - \mathbb{E}Z_2 > 1 - n^{-0.1}$$

This implies that a.a.s every pair of vertices $\{u, v\}$ is contained in at most two subhypergraphs R^{i} .

Finally, for $k \geq 3$, let

$$Z_k = \left| \left\{ S \in \binom{V'}{k} : Y_S \ge 2 \right\} \right|.$$

Then,

$$\mathbb{E}Z_k < n^k (n^{1.1})^2 (n^{-0.9})^{2k} = n^{k+2.2-1.8k} \le n^{-0.2}$$

Similarly,

$$\mathbb{P}(Z_k = 0) > 1 - n^{-0.2}$$

The latter implies that a.a.s. the induced subhypergraphs $H[R^i]$, $i = 1, ..., n^{1.1}$, are pairwise edge-disjoint. Summarizing, we can choose the sets R^i , $1 \le i \le n^{1.1}$ in such a way that

- (i) for every $v \in V', Y_{\{v\}} \sim n^{0.2}$,
- (ii) every pair $\{u, v\} \subset V'$ is contained in at most two sets R^i ,
- (iii) every edge $e \in H$ is contained in at most one set R^i ,
- (iv) for all $i = 1, \ldots, n^{1.1}$, we have $|R^i| \sim n^{0.1}$, and
- (v) for all $i = 1, ..., n^{1.1}$ and all $D \subseteq V', |D| = d$, we have $\text{DEG}_D^{(i)} \ge (c + \alpha/4) \binom{|R^i| d}{k d}$.

Let us fix a sequence R^i , $1 \le i \le n^{1.1}$, satisfying (i)-(v) above.
Our assumption that $f_d(k,n) \sim c^* \binom{n-d}{k-d}$ holds for all sufficiently large values of n, in particular with n replaced by $|R^i| \sim n^{0.1}$. Thus, we have

$$f_d(k, |R^i|) \sim c^* \binom{|R^i| - d}{k - d},$$

and, by condition (v) above, we conclude that

$$\delta_d(H[R^i]) \ge (c + \alpha/4) \binom{|R^i| - d}{k - d} > f_d(k, |R^i|).$$

Consequently, by the definition of f_d , there exists a fractional perfect matchings w^i in every subhypergraph $H[R^i]$, $i = 1, ..., n^{1.1}$.

Now comes the second round of randomization. Let $H^* = \bigcup_i H[R^i]$. We select a generalized binomial subhypergraph H'' of H^* by independently choosing each edge e with probability $w^{i_e}(e)$, where i_e is the index i such that $e \in H[R^i]$. Recall that property (iii) ensures that every edge is contained in at most one hypergraph R^i , which guarantees the uniqueness of i_e . We are going to verify our claim by showing $\deg_{H''}(v) \sim n^{0.2}$ for any vertex v, while $\Delta_2(H'') \leq n^{0.1}$.

Let $I_v = \{i : v \in \mathbb{R}^i\}$ and recall that $|I_v| = Y_{\{v\}} \sim n^{0.2}$ by (i). For every $v \in V'$ the set E_v of edges $e \in H^*$ containing v can be partitioned into $|I_v|$ parts $E_v^i = \{e \in E_v \cap H[\mathbb{R}^i]\}$. Recall that w^i is a perfect matching, and thus $\sum_{e \in E_v^i} w^i(e) = 1$. For every $v \in V'$ the random variable $D_v = \deg_{H''}(v)$ is equal to $\sum_{i \in I_v} \sum_{e \in E_v^i} X_e$, where X_e are independent random variables having Bernoulli distribution with expectation $w^{i_e}(e)$. Therefore D_v is generalized binomial with expectation

$$\mathbb{E}D_{v} = \sum_{e \in E_{v}} w^{i_{e}}(e) = \sum_{i \in I_{v}} \left(\sum_{e \in E_{v}^{i}} w^{i}(e)\right) = \sum_{i \in I_{v}} 1 \sim n^{0.2}.$$

Hence by Chernoff's inequality (4.4.1),

$$\mathbb{P}(|D_v - n^{0.2}| \ge \alpha n^{0.2}) \le 2e^{-\frac{\varepsilon^2}{3}n^{0.2}}$$

Set $\alpha = n^{-0.05}$, then $|D_v - n^{0.2}| \le n^{0.15}$ with probability $1 - O(e^{-n^{0.1}})$. Taking a union bound over all the *n* vertices, we conclude that a.a.s. for all $v \in V'$ we have $D_v \sim n^{0.2}$.

Moreover, for all pairs $u, v \in V'$ the random variable $D_{u,v} = \deg_{H''}(u, v)$ is also generalized binomial with expectation

$$\mathbb{E}D_{u,v} = \sum_{e \in E_u \cap E_v} w^{i_e}(e) = \sum_{i \in I_u \cap I_v} \left(\sum_{e \in E_u^i \cap E_v^i} w^i(e)\right) \le |I_u \cap I_v| \le 2$$

by (ii). Hence, again by Chernoff's inequality (4.4.3) for large deviations, when n is sufficiently large,

$$\mathbb{P}(D_{u,v} \ge n^{0.1}) \le e^{-n^{0.1}}$$

Once again taking the union bound ensures that a.a.s. for every pair of vertices $u, v \in V'$, $D_{u,v} \leq n^{0.1}$.

4.5 An application in distributed storage allocation

The following model of distributed storage has been studied in information theory [62, 74, 88]. A file is split into multiple chunks, replicated redundantly and stored in a distributed storage system with n nodes. Suppose the amount of data to be stored in each node i is equal to x_i , where the size of the whole file is normalized to 1. In reality, because there is limited storage space or transmission bandwidth, we require that the total amount of data stored does not exceed a given budget T, i.e. $x_1 + \cdots + x_n \leq T$. At the time of retrieval, we attempt to recover the whole file by accessing only the data stored in a subset R of r nodes which is chosen uniformly at random. It is known that there always exists a coding scheme such that we can recover the file whenever the total amount of data accessed is at least 1. Our goal is to find an optimal allocation (x_1, \dots, x_n) in order to maximize the probability of successful recovery. This problem can be reformulated as follows.

Question 4.5.1. For a sequence of nonnegative numbers (x_1, \dots, x_n) , let

$$\Phi(x_1, \cdots, x_n) = \Big| \big\{ S \subseteq [n], |S| = r \text{ such that } \sum_{i \in S} x_i \ge 1 \big\} \Big|.$$

Then the probability of successful recovery of the file equals

$$\frac{\Phi(x_1,\cdots,x_n)}{\binom{n}{r}}$$
64

Given integers $n \ge r \ge 1$ and a real number T > 0, determine

$$F^{T}(r,n) = \max_{\sum x_i = T, \ x_i \ge 0 \ \forall i} \Phi(x_1, \cdots, x_n).$$

and find an allocation optimizing $F^{T}(r, n)$.

In this section, we always assume that T is integer-valued in order to avoid any rounding issues. If the total budget T is at least n/r then, by setting all $x_i = T/n \ge 1/r$ for all i, we can recover the original file from any subset of size r. So, $F^T(r,n) = \binom{n}{r}$ for $T \ge n/r$. For T < n/r, let $w(i) = x_i$ be a weight function from V = [n] to \mathbb{R} . Then by the definition of the threshold r-uniform hypergraph H_w^1 from Section 4.3, the edges of H_w^1 correspond to the r-subsets S such that $\sum_{i \in S} x_i \ge 1$. Thus, it is easy to see that the fractional matching number of H^1_w satisfies

$$\nu^*(H^1_w) = \tau^*(H^1_w) \le \sum_{i=1}^n w(i) = \sum_{i=1}^n x_i \le T$$

while

$$\Phi(x_1,\cdots,x_n)=|H^1_w|.$$

Therefore, $F^{T}(r,n)$ is the maximum number of edges in an r-uniform hypergraph on n vertices with fractional matching number at most T. As such $F^T(r, n)$ differs from $f_0^T(r, n)$ only in that the latter has the strict inequality $\nu^*(H) < T$ in its definition. But, of course, we have $f_0^T(r, n) \le F^T(r, n) \le f_0^{T+1}(r, n)$, and so $F^T(r, n) \sim f_0^T(r, n)$ as $n \to \infty$.

Hence, Question 4.5.1 is asymptotically equivalent to the fractional Erdős Conjecture 4.1.9. As mentioned in the introduction, it follows from the Erdős-Gallai theorem [29] that

$$F^{T}(2,n) \sim f_{0}^{T}(2,n) \sim m_{0}^{T}(2,n) \sim \max\left\{ \binom{2T}{2}, \binom{n}{2} - \binom{n-T}{2} \right\}.$$

An easy calculation shows that the above maximum equals the first term if $\frac{2}{5}n \leq T \leq \frac{1}{2}n$, and the corresponding optimal graph is a clique of size 2T. This means that, asymptotically, an optimal allocation is $x_1 = \cdots = x_{2T} = 1/2$ and $x_{2T+1} = \cdots = x_n = 0$. On the other hand, if $T < \frac{2}{5}n$, an optimal allocation is $x_1 = \cdots = x_T = 1$ and $x_{T+1} = \cdots = x_n = 0$.

For general $r \geq 3$, if Conjecture 4.1.9 is true, then

$$F^{T}(r,n) \sim \max\left\{ \binom{rT}{r}, \binom{n}{r} - \binom{n-T}{r} \right\}.$$

65

The bounds are achieved when H is a clique or a complement of clique. A corresponding (asymptotically) optimal storage allocation is $x_1 = \cdots = x_{rT} = 1/r$, $x_{rT+1} = \cdots = x_n = 0$ or $x_1 = \cdots = x_T = 1$, $x_{T+1} = \cdots = x_n = 0$, respectively. Corollary 4.2.5 and Remark 4.2.6 assert that for r = 3 and T < 0.277 n, as well as for r = 4 and T < 0.217 n, the latter is an optimal allocation. Moreover, if Samuels' conjecture 4.2.1 holds for all the remaining $r \ge 5$, then $x_1 = \cdots = x_T = 1$, $x_{T+1} = \cdots = x_n = 0$ is always an asymptotic optimal allocation whenever T < n/(r+1). Erdős [28] proved Conjecture 4.1.8 for all $T < n/(2r^3)$. In Chapter 3, we extended the range for which this conjecture holds to $T = O(n/r^2)$. Therefore, in this range, $F^T(r, n)$ is achieved by the complement of a clique and an optimal allocation is also known to be $x_1 = \cdots = x_T = 1$, $x_{T+1} = \cdots = x_n = 0$.

4.6 Concluding Remarks

In this chapter, we have studied sufficient conditions on the minimum *d*-degree which guarantee that a uniform hypergraph has a perfect matching or perfect fractional matching. We proved that if $f_d(k,n) \sim c^*\binom{n}{k}$, then $m_d(k,n) \sim \max\{c^*, 1/2\}\binom{n}{k}$. Therefore in order to determine the asymptotic behavior of the minimum *d*-degree ensuring existence of a perfect matching, we can instead study the presumably easier question for fractional matchings. Using this approach we showed, in particular, that $m_1(5,n) \sim \left(1 - \frac{4^4}{5^4}\right)\binom{n-1}{4}$.

An intriguing problem which remains open is the conjecture by Erdős which states that the maximum number of edges in a k-uniform hypergraph H on n vertices with matching number smaller than s is exactly

$$\max\left\{\binom{ks-1}{k}, \binom{n}{k} - \binom{n-s+1}{k}\right\}.$$

The fractional version of Erdős conjecture is also very interesting. In its asymptotic form it says that if H is an l-uniform m-vertex hypergraph with fractional matching number $\nu^*(H) = xm$, where $0 \le x < 1/l$, then

$$|H| \le (1+o(1)) \max\left\{ (lx)^l, 1-(1-x)^l \right\} \binom{m}{l}.$$

In Section 4.2 we showed that the fractional Erdős conjecture is related to a probabilistic

conjecture of Samuels. This conjecture, if proved, will provide a solution to the fractional version of Erdős problem for the range $x \leq \frac{1}{l+1}$. It will also lead to the asymptotics of $m_d(k,n)$ and $f_d(k,n)$ for arbitrary $k \geq d+1$ and $d \geq 1$.

As it turns out, matchings and fractional matchings also have some interesting applications in information theory. In particular, the uniform model of distributed storage allocation considered in [88] leads to a question which is asymptotically equivalent to the fractional version of Erdős' problem. In [62], the set of accessed nodes, R, is given by taking each node randomly and independently with probability p. It would be interesting to see if our techniques can be applied to study this binomial model too.

Acknowledgments The author would like to thank Alex Dimakis for the discussion on the fractional Erdős conjecture.

CHAPTER 5

Extremal problems in Eulerian digraphs

5.1 Introduction

One of the central themes in graph theory is to study the extremal graphs which satisfy certain properties. Extremity can be taken with respect to different parameters as order, size, or girth. There are many classical results in this area. For example, any undirected graph G with n vertices and m edges has a subgraph with minimum degree at least m/n, and thus G also contains a cycle of length at least m/n + 1. It is natural to ask whether such results can be extended to digraphs, in which every edge is associated with a direction. However, it turns out that these statements are often trivially false even for very dense general digraphs. For instance, a transitive tournament does not contain any cycle, and its subgraphs always have zero minimum in-degree and out-degree. Therefore in order to obtain meaningful results as in the undirected case, it is necessary to restrict to a smaller family of digraphs. A natural candidate one may consider is the family of *Eulerian digraphs*, in which the in-degree equals the out-degree at each vertex. In this chapter, we investigate several natural parameters of Eulerian digraphs, and study the connections between them. In particular, the parameters we consider are minimum feedback arc set, shortest cycle, longest cycle, and largest minimum degree subgraph. Throughout this chapter, we always assume the Eulerian digraph is simple, i.e. it has no multiple arcs or loops, but arcs in different directions like (u, v) and (v, u) are allowed. For other standard graph-theoretic terminology involved, the reader is referred to [13].

A feedback arc set of a digraph is a set of arcs whose removal makes the digraph acyclic. Given a digraph G, denote by $\beta(G)$ the minimum size of a feedback arc set. Computing $\beta(G)$ and finding a corresponding minimum feedback arc set is a fundamental problem in combinatorial optimization. It has applications in many other fields such as testing of electronic circuits and efficient deadlock resolution (see, e.g., [61, 91]). However, computing $\beta(G)$ turns out to be difficult, and it is NP-hard even for tournaments [3, 17]. One basic question in this area is to bound $\beta(G)$ as a function of other parameters of G, and there are several papers (see, e.g., [18, 33, 95]) studying upper bounds for $\beta(G)$ of this form. However, much less is known for the lower bound of $\beta(G)$, perhaps because a general digraph could be very dense and still have a small minimum feedback arc set. For example, a transitive tournament has $\beta(G) = 0$. Nevertheless, it is easy to see that any Eulerian digraph G with n vertices and m arcs has $\beta(G) \geq m/n$, since the arcs can be decomposed into a disjoint union of cycles, each of length at most n, and any feedback arc set contains at least one arc from each cycle. In this chapter we actually prove the following much stronger lower bound for $\beta(G)$.

Theorem 5.1.1. Every Eulerian digraph G with n vertices and m arcs has $\beta(G) \ge m^2/2n^2 + m/2n$.

Moreover, Theorem 5.1.1 is tight for an infinite family of Eulerian digraphs, as can be seen from the following proposition.

Proposition 5.1.2. For every pair of integers m and n such that m is divisible by n, there exists an Eulerian digraph G with n vertices and m arcs, and with $\beta(G) = m^2/2n^2 + m/2n$.

The study of the existence of cycles plays a very important role in graph theory, and there are numerous results for undirected graphs in the classical literature. However, there are significantly fewer results for digraphs. The main reason for this is probably because digraphs often behave more similar to hypergraphs, and questions concerning cycles in digraphs are often much more difficult than the corresponding questions in graphs. One of the most famous problems in this area is the celebrated Caccetta-Häggkvist conjecture [16]: every directed *n*-vertex digraph with minimum outdegree at least *r* contains a cycle with length at most $\lceil n/r \rceil$, which is not completely solved even when restricted to Eulerian digraphs (for more discussion, we direct the interested reader to the surveys [75, 96]). In this chapter, we study the existence of short cycles in Eulerian digraphs with a given order and size. The girth g(G) of a digraph G is defined as the length of the shortest cycle in G. Combining Theorem 5.1.1 and a result of Fox, Keevash and Sudakov [33] which connects $\beta(G)$ and g(G) for a general digraph G, we are able to obtain the following corollary.

Corollary 5.1.3. Every Eulerian digraph G with n vertices and m arcs has $g(G) \leq 6n^2/m$.

We also point out that the upper bound in Corollary 5.1.3 is tight up to a constant, since the construction of Proposition 5.1.2 also provides an example of Eulerian digraphs with girth at least n^2/m .

A repeated application of Corollary 5.1.3 gives an Eulerian subgraph of the original digraph G, whose arc set is a disjoint union of $\Omega(m^2/n^2)$ cycles. Using this fact we can find an Eulerian subgraph of G with large minimum degree.

Theorem 5.1.4. Every Eulerian digraph G with n vertices and m arcs has an Eulerian subgraph with minimum degree at least $m^2/24n^3$. This bound is tight up to a constant for infinitely many m, n.

In 1996, Bollobás and Scott ([15], Conjecture 6) asked whether every Eulerian digraph G with nonnegative arc-weighting w contains a cycle of weight at least cw(G)/(n-1), where w(G) is the total weight and c is some absolute constant. For the unweighted case, i.e. w = 1, this conjecture becomes: "Is it true that every Eulerian digraph with n vertices and m arcs contains a cycle of length at least cm/n?" Even this special case is still wide open after 15 years. An obvious consequence of Theorem 5.1.4 is that every Eulerian digraph contains a cycle of length at least $1 + m^2/24n^3$. When the digraph is dense, i.e. $m = cn^2$, our theorem provides a cycle of length linear in n, which partially verifies the Bollobás-Scott conjecture in this range. However observe that when m is small, in particular when $m = o(n^{3/2})$, Theorem 5.1.4 becomes meaningless. Nevertheless, we can always find a long cycle of length at least $\lfloor \sqrt{m/n} \rfloor + 1$, as shown by the following proposition¹.

¹This proposition was also obtained independently by Jacques Verstraete.

Proposition 5.1.5. Every Eulerian digraph G with n vertices and m arcs has a cycle of length at least $1 + \lfloor \sqrt{m/n} \rfloor$. Together with Theorem 5.1.4, this implies that G has a cycle of length at least $1 + \max\{m^2/24n^3, \lfloor \sqrt{m/n} \rfloor\}$.

The rest of this chapter is organized as follows. In Section 5.2, we obtain our bounds for feedback arc sets by proving Theorem 5.1.1 and Proposition 5.1.2. Section 5.3 contains the proof of our results for the existences of short cycles, long cycles, and subgraph with large minimum degree. The final section contains some concluding remarks and open problems.

5.2 Feedback arc sets

This section contains the proofs of Theorem 5.1.1 and Proposition 5.1.2. Consider some linear order of the vertex set of an Eulerian digraph G = (V, E) with *n* vertices and *m* arcs. Let v_i is the *i*'th vertex in this order. We say that v_i is before v_j if i < j. An arc (v_i, v_j) is a forward arc if i < j, and is a backward arc if i > j. Observe that any cycle contains at least one backward arc. Hence, the set of backward arcs forms a feedback arc set. We prove Theorem 5.1.1 by showing that any linear order of V has at least as many backward arcs as the amount stated in the theorem. We first require the following simple lemma. Here a *cut* is defined as a partition of the vertices of a digraph into two disjoint subsets.

Lemma 5.2.1. In any cut $(A, V \setminus A)$ of an Eulerian digraph, the number of arcs from A to $V \setminus A$ equals the number of arcs from $V \setminus A$ to A.

Proof. The sum of the out-degrees of the vertices of A equals the sum of the in-degrees of the vertices of A. Each arc with both endpoints in A contributes one unit to each of these sums. Hence, the number of arcs with only one endpoint in A splits equally between arcs that go from A to $V \setminus A$ and arcs that go from $V \setminus A$ to A.

Proof of Theorem 5.1.1. Fix an Eulerian digraph G with |V| = n and |E| = m. We claim that it suffices to only consider Eulerian digraphs which are 2-cycle-free, i.e. between any pair of vertices $\{i, j\}$, there do not exist arcs in two different directions. Suppose there

are k different 2-cycles in G. By removing all of them, we delete exactly 2k arcs. Note that the resulting 2-cycle-free digraph G' is still Eulerian and contains m - 2k arcs. Therefore if Theorem 5.1.1 is true for all 2-cycle-free Eulerian digraphs, then

$$\beta(G') \geq \frac{(m-2k)^2}{2n^2} + \frac{m-2k}{2n} \,.$$

Obviously in any linear order of V(G), exactly half of the 2k arcs deleted must be backward arcs. Therefore,

$$\begin{split} \beta(G) &\geq \beta(G') + k \geq \frac{(m-2k)^2}{2n^2} + \frac{m-2k}{2n} + k = \left(\frac{m^2}{2n^2} + \frac{m}{2n}\right) - \frac{2k(m-k)}{n^2} + k - \frac{k}{n} \\ &\geq \left(\frac{m^2}{2n^2} + \frac{m}{2n}\right) - \frac{2k\binom{n}{2}}{n^2} + k - \frac{k}{n} = \frac{m^2}{2n^2} + \frac{m}{2n} \;. \end{split}$$

The last inequality follows from the fact that $m - k \leq {n \choose 2}$, since m - k counts the number of pairs of vertices with an arc between them.

From now on, we always assume that G is a 2-cycle-free Eulerian digraph. In order to prove a lower bound on $\beta(G)$, we fix a linear order v_1, \ldots, v_n . It will be important for the analysis to consider the *length* of an arc (v_i, v_j) which is |i - j|. Observe that the length of any arc is an integer in $\{1, \ldots, n-1\}$. Moreover, we call an arc *short* if its length is at most n/2. Otherwise, it is *long*.

Partition the arc set E into two parts, S and L, where S contains the short arcs and L contains the long arcs. For a vertex v_i , let s_i denote the number of short arcs connecting v_i with some v_j where j > i. It is important to note that at this point we claim nothing regarding the directions of these arcs. Since G is 2-cycle-free, $s_i \leq n - i$. As each short arc (v_i, v_j) contributes exactly one to either s_i or s_j , we have that:

$$\sum_{i=1}^{n} s_i = |S|$$

We now estimate the sum of the lengths of the short arcs. Consider some vertex v_i . Since G is 2-cycle-free, the s_i short arcs connecting v_i to vertices appearing after v_i must have distinct lengths. Hence, the sum of their lengths is at least $1 + 2 + \cdots + s_i = {s_i+1 \choose 2}$. Thus, denoting by w(S) the sum of the lengths of the short arcs, we have that

$$w(S) \ge \sum_{i=1}^{n} {\binom{s_i+1}{2}}.$$
 (5.2.1)

Next we calculate the sum of the lengths of the long arcs, that is denoted by w(L). There is at most one long arc of length n - 1. There are at most two arcs of length n - 2, and, more generally, there are at most n - i arcs of length i. Thus, if we denote by t_i the number of long arcs of length i for $i \ge \lfloor n/2 \rfloor + 1$ and set $t_i = 0$ for $i \le \lfloor n/2 \rfloor$, we have that $t_i \le n - i$, and

$$w(L) = \sum_{i=1}^{n} i \cdot t_i .$$
 (5.2.2)

Obviously,

$$\sum_{i=1}^{n} t_i + \sum_{i=1}^{n} s_i = |L| + |S| = m \; .$$

Let $A_i = \{v_1, \ldots, v_i\}$ and consider the cuts $C_i = (A_i, V \setminus A_i)$ for $i = 1, \ldots, n$. Let c_i denote the number of arcs crossing C_i (and notice that $c_n = 0$). Since an arc of length x crosses precisely x of these cuts, we have that

$$\sum_{i=1}^{n} c_i = w(S) + w(L) .$$
(5.2.3)

Consider a pair of cuts $C_i, C_{i+\lfloor n/2 \rfloor}$ for $i = 1, \ldots, \lfloor n/2 \rfloor$. If an arc crosses both C_i and $C_{i+\lfloor n/2 \rfloor}$ then its length is at least $\lfloor n/2 \rfloor + 1$. Hence, a short arc cannot cross both of these cuts. Let y_i denote the number of long arcs that cross both of these cuts. By Lemma 5.2.1, $c_i/2$ backward arcs cross C_i and $c_{i+\lfloor n/2 \rfloor}/2$ backward arcs cross $C_{i+\lfloor n/2 \rfloor}$, and we have counted at most y_i such arcs twice. It follows that the number of backward arcs is at least

$$\frac{1}{2}(c_i+c_{i+\lfloor n/2\rfloor})-y_i.$$

Averaging over all $\lfloor n/2 \rfloor$ such pairs of cuts, it follows that the number of backward arcs is at least

$$\frac{1}{\lfloor n/2 \rfloor} \sum_{i=1}^{\lfloor n/2 \rfloor} \left(\frac{1}{2} (c_i + c_{i+\lfloor n/2 \rfloor}) - y_i \right) .$$
(5.2.4)

As each long arc of length j crosses precisely $j - \lfloor n/2 \rfloor$ pairs of cuts C_i and $C_{i+\lfloor n/2 \rfloor}$, we have that $\sum_{i=1}^{\lfloor n/2 \rfloor} y_i = \sum_{j \ge \lfloor n/2 \rfloor} t_j (j - \lfloor n/2 \rfloor) = w(L) - |L| \cdot \lfloor n/2 \rfloor$. This, together with (5.2.3) and (5.2.4) gives that

$$\beta(G) \geq \frac{1}{\lfloor n/2 \rfloor} \left(\frac{1}{2} (w(S) + w(L)) - (w(L) - |L| \cdot \lfloor n/2 \rfloor) \right)$$

$$\geq \frac{w(S) - w(L)}{2 \lfloor n/2 \rfloor} + |L|. \qquad (5.2.5)$$

Note that when n = 2k is even, the above inequality becomes

$$\beta(G) \ge \frac{w(S) - w(L)}{n} + |L|$$

Next we show that when n = 2k + 1 is odd, the same inequality still holds. To see this, first assume that $w(S) \ge w(L)$. Then applying inequality (5.2.5), we have that for n = 2k + 1,

$$\beta(G) \ge \frac{w(S) - w(L)}{2k} + |L| \ge \frac{w(S) - w(L)}{n} + L$$
.

Next suppose that w(S) < w(L). Instead of considering the cuts C_i and C_{i+k} , we look at the pair C_i and C_{i+k+1} for $i = 1, \dots, k$. Moreover, denote by z_i the number of long arcs that cross both of these cuts. By a similar argument as before, the number of backward arcs is at least $\frac{1}{2}(c_i + c_{i+k+1}) - z_i$ for $1 \le i \le k$, and $c_i/2$ for i = k + 1. This provides k + 1 lower bounds for $\beta(G)$, and we will average over all of them. Since each long arc of length j crosses precisely j - (k + 1) pairs of cuts C_i and C_{i+k+1} , we again have that $\sum_{i=1}^{k} z_i = \sum_{j \ge k+1} t_j (j - (k + 1)) = w(L) - (k + 1) |L|$, and we have that

$$\beta(G) \geq \frac{1}{k+1} \left(\sum_{i=1}^{k} \left(\frac{1}{2} (c_i + c_{i+k+1}) - z_i \right) + \frac{c_{k+1}}{2} \right)$$

$$\geq \frac{1}{k+1} \left(\frac{1}{2} (w(S) + w(L)) - (w(L) - (k+1)|L|) \right)$$

$$\geq \frac{w(S) - w(L)}{2k+2} + |L| \geq \frac{w(S) - w(L)}{n} + |L|,$$

where we use the fact that w(L) > w(S).

Using our lower bound estimate (5.2.1) for w(S) and the expression (5.2.2) for w(L), we obtain that

$$\beta(G) \geq \frac{w(S) - w(L)}{n} + |L|
\geq \frac{1}{n} \left(\sum_{i=1}^{n} {s_i + 1 \choose 2} - \sum_{i=1}^{n} i \cdot t_i \right) + \sum_{i=1}^{n} t_i$$

$$= \frac{1}{n} \left(\sum_{i=1}^{n} {s_i + 1 \choose 2} + (n - i)t_i \right).$$
(5.2.6)

Define

$$F(s_1, \cdots, s_n; t_1, \cdots, t_n) := \sum_{i=1}^n {\binom{s_i+1}{2}} + (n-i)t_i$$

In order to find a lower bound of $\beta(G)$, we need to solve the following integer optimization problem.

$$F(m,n) := \min F(s_1, \cdots, s_n; t_1, \cdots, t_n)$$

subject to $s_i \le n-i$, $t_i \le n-i$, $\sum_{i=1}^n s_i + \sum_{i=1}^n t_i = m$.

The following Lemma 5.2.2 provides a precise solution to this optimization problem, which gives that $F(m,n) = tm - (t^2 - t)n/2$, where $t = \lceil m/n \rceil$. Hence if we assume that m = tn - k with $0 \le k \le n - 1$, then

$$\begin{split} \beta(G) &\geq \frac{1}{n} F(m,n) = \frac{tm}{n} - \frac{t^2 - t}{2} = \frac{t(tn-k)}{n} - \frac{t^2 - t}{2} \\ &= \frac{t^2 + t}{2} - \frac{tk}{n} \geq \frac{t^2 + t}{2} - \frac{tk}{n} + \left(\frac{k^2}{2n^2} - \frac{k}{2n}\right) \\ &= \frac{(tn-k)^2}{2n^2} + \frac{tn-k}{2n} = \frac{m^2}{2n^2} + \frac{m}{2n} \;. \end{split}$$

The last inequality is because $0 \le k \le n-1$, so $0 \le k/n < 1$ and $k^2/2n^2 \le k/2n$. Note that equality is possible only when m is a multiple of n.

Lemma 5.2.2. $F(m,n) = tm - (t^2 - t)n/2$, where $t = \lceil m/n \rceil$.

Proof. The proof of this lemma consists of several claims. We assume that $s_i + t_i = a_i$, then $0 \le a_i \le 2(n-i)$ and $s_i \le n-i$, so

$$\binom{s_i+1}{2} + (n-i)t_i = \frac{1}{2}s_i^2 - (n-i-1/2)s_i + (n-i)a_i.$$

Since s_i is an integer, this function of s_i is minimized when $s_i = n - i$ if $a_i \ge n - i$, and when $s_i = a_i$ if $a_i < n - i$. Therefore, subject to $\sum_i a_i = m$ and $a_i \le 2(n - i)$, we want to minimize

$$F = \sum_{a_i < n-i} \binom{a_i + 1}{2} + \sum_{a_i \ge n-i} \left(\binom{n-i+1}{2} + (n-i)(a_i - (n-i)) \right)$$
$$= \sum_{a_i < n-i} \binom{a_i + 1}{2} + \sum_{a_i \ge n-i} \left((n-i)a_i - \binom{n-i}{2} \right).$$
(5.2.7)

For convenience, define $A = \{i : a_i < n - i\}$, and $B = \{i : a_i \ge n - i\}$.

Claim 1. For any $i \in A$, if we increase a_i by 1 then F increases by $a_i + 1$, and if we decrease a_i by 1 then F decreases by a_i . For any $j \in B$, if we increase (decrease) a_j by 1 then F increases (decreases) by n - j.

Proof. Note that when $a_i = n - i$ or n - i - 1, $\binom{a_i + 1}{2} = (n - i)a_i - \binom{n - i}{2}$, therefore if we increase a_i by 1 for any $i \in A$, the contribution of a_i to F always increases by $\binom{a_i + 2}{2} - \binom{a_i + 1}{2} = a_i + 1$. When we decrease a_i by 1, F decreases by $\binom{a_i + 1}{2} - \binom{a_i}{2} = a_i$. It is also easy to see that for any $j \in B$, if we increase or decrease a_j by 1, the contribution of a_j to F always increases or decreases by n - j.

Next we show that for any extremal configuration (a_1, \dots, a_n) which minimizes F, any integer of A is smaller than any integer of B.

Claim 2. F is minimized when $A = \{1, \dots, l-1\}$ and $B = \{l, \dots, n\}$ for some integer l.

Proof. We prove by contradiction. Suppose this statement is false, then F is minimized by some $\{a_i\}_{i=1}^n$ such that there exists i < j, $i \in B$ and $j \in A$. Now we decrease a_i by 1 and increase a_j by 1, which can be done since $a_j < 2(n-j)$. Then by Claim 1, F decreases by $(n-i) - (a_j + 1) \ge n - (j-1) - (a_j + 1) = (n-j) - a_j > 0$ since $j \in A$, which contradicts the minimality of F.

Since $\sum_{i=1}^{n} a_i = m$, which is fixed. The next claim shows that in order to minimize F, we need to take the variables whose index is in B to be as large as possible, with at most one exception.

Claim 3. F is minimized when $A = \{1, \dots, l-1\}$, and $B = \{l, \dots, n\}$ for some integer l. Moreover, $a_i = 2(n-i)$ for all $i \ge l+1$.

Proof. First note that for $i \in B$, its contribution to F is $(n-i)a_i - \binom{n-i}{2}$. The second term is fixed, and a_i has coefficient n-i which decreases in i. Therefore when F is minimized, if iis the largest index in B such that $a_i < 2(n-i)$, then all j < i in B must satisfy $a_j = n - j$; otherwise we might decrease a_j and increase a_i to make F smaller. Therefore, if i > l, we have $a_{i-1} = n - i + 1$. Note that if we increase a_i by 1 and decrease a_{i-1} by 1, by Claim 1 the target function F decreases by $a_{i-1} - (n-i) = 1$. Therefore the only possibility is that i = l, which proves Claim 3.

Claim 4. There is an extremal configuration for which $a_i = n - l$ or $a_i = n - l + 1$ for $i \leq l - 1$, a_l is between n - l and 2(n - l), and $a_i = 2(n - i)$ for $i \geq l + 1$.

Proof. From Claim 3, we know that in an extremal configuration, $a_i < n-i$ for $1 \le i \le l-1$, $n-l \le a_l \le 2(n-l)$, and $a_i = 2(n-i)$ for $i \ge l+1$. Among all extremal configurations, we take one with the largest l, and for all such configurations, we take one for which a_l is the smallest. For such a configuration, if we increase a_j by 1 for some $j \in A$ and decrease a_l by 1, then by Claim 1, F increases by $(a_j + 1) - (n - l)$, which must be nonnegative. Suppose $a_j + 1 = n - l$. If j is changed to be in B, it contradicts Claim 3 no matter whether l remains in B or is changed to be in A; if j remains in A, it contradicts the maximality of l if l is changed to be in A or contradicts the minimality of a_l if l remains in B. Therefore $a_j \ge n - l$ for every $1 \le j \le l - 1$. We next consider two cases: either a_l is equal to 2(n-l), or strictly less than 2(n-l).

<u>Case 1.</u> $a_l = 2(n-l)$. From the discussions above, we already know that $a_j \ge n-l$ for every $1 \le j \le l-1$. In particular $a_{l-1} = n-l$ since it is strictly less than n - (l-1). If for some $j \le l-1$, $a_j \ge n-l+2$, then we can decrease a_j by 1 and increase a_{l-1} by 1 since a_j is strictly greater than 0 and a_{l-1} is strictly less than 2(n-l+1). By Claim 1, Fdecreases by $a_j - (n-l+1) \ge 1$, which contradicts the minimality of F. Hence we have that $n-l \le a_j \le n-l+1$ for every $j \le l-1$.

<u>Case 2.</u> $a_l < 2(n-l)$. If we decrease a_j by 1 and increase a_l by 1, F decreases by $a_j - (n-l)$ by Claim 1, therefore $a_j \le n-l$ by the minimality of F, hence $a_j = n-l$ for all $1 \le j \le l-1$.

In both cases, the extremal configuration consists of n - l or n - l + 1 for the first l - 1 variables, a_l is between n - l and 2(n - l), and $a_i = 2(n - i)$ for $i \ge l + 1$.

By Claim 4, we can bound the number of arcs m from both sides,

$$m = \sum_{i=1}^{l-1} a_i + \sum_{i=l}^n a_i \ge (l-1)(n-l) + (n-l) + \sum_{i=l+1}^n 2(n-i) = (n-l)(n-1) .$$

$$m = \sum_{i=1}^{l-1} a_i + \sum_{i=l}^n a_i < (l-1)(n-l+1) + \sum_{i=l}^n 2(n-i) = (n-l+1)(n-1) .$$

Solving these two inequalities, we get

$$n - \frac{m}{n-1} \le l < n+1 - \frac{m}{n-1}$$

Let m = tn - k, where $t = \lceil m/n \rceil$ and $0 \le k \le n - 1$. It is not difficult to check that if $t \ge k, l = n - t$ and if t < k, l = n - t + 1.

Now let x be the number of variables $a_1, ..., a_{l-1}$ which are equal to n - l + 1. Since $a_i = 2(n-i)$ for $i \ge l+1$, we have that

$$x + a_l = m - (l - 1)(n - l) - \sum_{i \ge l+1} a_i = m - (n - 2)(n - l) .$$
 (5.2.8)

When $t \ge k$, then l = n - t and

$$x + a_l = m - (n - 2)t = 2t - k < 2t = 2(n - l),$$

hence $a_l < 2(n-l)$. By the analysis of the second case in Claim 4, $a_j = n - l = t$ for all $j \le l-1$, therefore x = 0 and $a_l = 2t - k$. Since l = n - t, then using the summation formula $\sum_{k=1}^{n} k^2 = k(k+1)(2k+1)/6$, we have from (5.2.7) that (with details of the calculation omitted)

$$F = \binom{t+1}{2}(n-t-1) + t(2t-k) - \binom{t}{2} + \sum_{i \ge l+1} \left(2(n-i)^2 - \binom{n-i}{2}\right) = tm - (t^2 - t)n/2.$$

Now we assume t < k, then l = n - t + 1. Then using (5.2.8) again,

 $x + a_l = m - (n - 2)(t - 1) = n - k + 2(t - 1) > 2(t - 1) = 2(n - l) .$

The only possibility without contradicting the second case in Claim 4 is that $a_l = 2(n - l)$ and x = n - k. Thus, there are n - k of $a_1, ..., a_{l-1}$ which are equal to n - l + 1 = t and the rest k - t are equal to t - 1. Again by (5.2.7),

$$F = \binom{t+1}{2}(n-k) + \binom{t}{2}(k-t) + \sum_{i \ge l} \left(2(n-i)^2 - \binom{n-i}{2}\right) = tm - (t^2 - t)n/2.$$

As we have covered both cases, we have completed the proof of Lemma 5.2.2.

Proof of Proposition 5.1.2. Now we construct an infinite family of Eulerian digraphs which achieve the bound in Theorem 5.1.1. For any positive integers n, m such that t := m/n is an integer, we define the Cayley digraph G(n, m) to have vertex set $\{1, 2, ..., n\}$ and arc set $\{(i, i + j) : 1 \le i \le n, 1 \le j \le t\}$, where all additions are modulo n. From the definition, it is easy to verify that G(n, m) is an Eulerian digraph. Consider an order of the vertex set such that vertex i is the i'th vertex in this order, we observe that for $n - t + 1 \le i \le n$, vertex i has backward arcs (i, j), where $1 + n - i \le j \le t$ and there is no backward arc from vertex i for $i \le n - t$. Therefore,

$$\beta(G(n,m)) \le \sum_{i=n-t+1}^{n} t - (n-i) = \sum_{j=1}^{t} j = \binom{t+1}{2} = \frac{m^2}{2n^2} + \frac{m}{2n} .$$

5.3 Short cycles, long cycles, and Eulerian subgraphs with high minimum degree

In this section, we prove the existence of short cycles, long cycles, and subgraphs with large minimum degree in Eulerian digraphs. An important component in our proofs is the following result by Fox, Keevash and Sudakov [33] on general digraphs. We point out that the original Theorem 1.2 in [33] was proved with a constant 25, which can be improved to 18 using the exact same proof if we further assume $r \ge 11$.

Theorem 5.3.1. If a digraph G with n vertices and m arcs has $\beta(G) > 18n^2/r^2$, with $r \ge 11$, then G contains a cycle of length at most r, i.e. $g(G) \le r$.

Applying this theorem and Theorem 5.1.1, we can now prove Corollary 5.1.3, which says that every Eulerian digraph G with n vertices and m arcs contains a cycle of length at most $6n^2/m$.

Proof of Corollary 5.1.3. Given an Eulerian digraph G with n vertices and m arcs, if G contains a 2-cycle, then $g(G) \le 2 \le 6n^2/m$. So we may assume that G is 2-cycle-free and

thus $m \leq \binom{n}{2}$. By Theorem 5.1.1,

$$\beta(G) \ge \frac{m^2}{2n^2} + \frac{m}{2n} > \frac{m^2}{2n^2} = \frac{18n^2}{(6n^2/m)^2}$$

Since $r = 6n^2/m > 6n^2/\binom{n}{2} > 11$, we can use Theorem 5.3.1 to conclude that

$$g(G) \le r = \frac{6n^2}{m}$$

To see that this bound is tight up to a constant factor, we consider the construction of the Cayley digraphs in Proposition 5.1.2. It is not hard to see that if k = m/n, the shortest directed cycle in G(n,m) has length at least $\lceil n/k \rceil \ge n^2/m$.

Next we show that every Eulerian digraph with n vertices and m arcs has an Eulerian subgraph with minimum degree $\Omega(m^2/n^3)$.

Proof of Theorem 5.1.4. We start with an Eulerian digraph G with n vertices and m arcs. Note that Corollary 5.1.3 implies that every Eulerian digraph with n vertices and at least m/2 arcs contains a cycle of length at most $12n^2/m$. In every step, we pick one such cycle and delete all of its arcs from G. Obviously the resulting digraph is still Eulerian, and this process will continue until there are less than m/2 arcs left in the digraph. Therefore through this process we obtain a collection \mathscr{C} of t arc-disjoint cycles C_1, \dots, C_t , where $t \geq (m - m/2)/(12n^2/m) \geq m^2/24n^2$. Denote by H the union of all these cycles, obviously H is an Eulerian subgraph of G.

If H has minimum degree at least $\lceil t/n \rceil \ge m^2/24n^3$, then we are already done. Otherwise, we repeatedly delete from H any vertex v with degree $d(v) \le \lceil t/n \rceil - 1$, together with all the d(v) cycles in \mathscr{C} passing through v. This process stops after a finite number of steps. In the end we delete at most $n(\lceil t/n \rceil - 1) \le t - 1$ cycles in \mathscr{C} , so the resulting digraph H'is nonempty. Moreover, every vertex in H' has degree at least $\lceil t/n \rceil \ge m^2/24n^3$. Since H'is the disjoint union of the remaining cycles, it is also an Eulerian subgraph of G, and we conclude the proof of Theorem 5.1.4.

Remark. The proof of Theorem 5.1.4 also shows that G contains an Eulerian subgraph with minimum degree $\Omega(m^2/n^3)$ and at least $\Omega(m)$ arcs.

To see that the bound in Theorem 5.1.4 is tight up to a constant, for any integers s, t > 0, we construct an Eulerian digraph H := H(s, t) such that

- $V(H) = (U_1 \cup \dots \cup U_s) \cup (V_1 \cup \dots \cup V_t), |U_i| = |V_j| = s \text{ for } 1 \le i \le s, 1 \le j \le t,$
- for any $1 \le i \le t 1$ and vertices $u \in V_i, v \in V_{i+1}$, the arc $(u, v) \in E(H)$,
- for any $1 \le i \le s$ and every vertex $u \in U_i$, there is an arc from u to the *i*'th vertex in V_1 , and another arc from the *i*'th vertex in V_t to u.



Figure 5.1: The Eulerian digraph H(s,t) with s = 3

It can be verified that H(s,t) is an Eulerian digraph with (s+t)s vertices and $s^2(t+1)$ arcs. Moreover, every cycle in H(s,t) must pass through a vertex in $U_1 \cup \cdots \cup U_s$, whose degree is exactly 1. Therefore any Eulerian subgraph of H(s,t) has minimum degree at most 1. Next we define the δ -blowup $H(s,t,\delta)$: for any integer $\delta > 0$, we replace every vertex $i \in V(H(s,t))$ with an independent set $|W_i| = \delta$, and each arc $(i,j) \in E(H(s,t))$ by a complete bipartite digraph with arcs directed from W_i to W_j . The blowup digraph $H(s,t,\delta)$ is still Eulerian, and has $n = s(s+t)\delta$ vertices and $m = s^2(t+1)\delta^2$ arcs. Taking t = 2s, we have that for $H(s, 2s, \delta)$,

$$\frac{m^2}{n^3} = \frac{(s^2(2s+1)\delta^2)^2}{(s(s+2s)\delta)^3} = \frac{1}{27}\left(2+\frac{1}{s}\right)^2\delta \ge \frac{4}{27}\delta.$$

Note that similarly with the previous discussion on H(s,t), every cycle in the blowup $H(s, 2s, \delta)$ contains at least one vertex with degree δ . Therefore, the minimum degree of any Eulerian subgraph of $H(s, 2s, \delta)$ is at most $\delta \leq \frac{27}{4} \frac{m^2}{n^3}$. This implies that the bound in Theorem 5.1.4 is tight up to a constant factor for infinitely many m, n.

Before proving Proposition 5.1.5, let us recall the following easy fact.

Proposition 5.3.2. If a digraph G has minimum outdegree $\delta^+(G)$, then G contains a directed cycle of length at least $\delta^+(G) + 1$.

Proof. Let $P = v_1 \rightarrow v_2 \rightarrow \cdots \rightarrow v_t$ be the longest directed path in G. Then all the out neighbors of v_t must lie on this path, otherwise P will become longer. If i < t is minimal with $(v_t, v_i) \in E(G)$, then $v_i \rightarrow \cdots \rightarrow v_t \rightarrow v_i$ gives a cycle of length at least $d^+(v_t) + 1 \geq \delta^+(G) + 1$.

This proposition, together with Theorem 5.1.4, shows that an Eulerian digraph G with n vertices and m arcs contains a cycle of length at least $1 + m^2/24n^3$. But as we discussed in the introduction, this bound becomes meaningless when the number of arcs m is small. However, we may use a different approach to obtain a cycle of length at least $\lfloor \sqrt{m/n} \rfloor + 1$.

Proof of Proposition 5.1.5. To prove that any Eulerian digraph G with n vertices and m arcs has a cycle of length at least $\lfloor \sqrt{m/n} \rfloor + 1$, we use induction on the number of vertices n. Note that the base case when n = 2 is obvious, since the only Eulerian digraph is the 2-cycle with $\lfloor \sqrt{m/n} \rfloor + 1 = 2$. Suppose the statement is true for n-1. Consider an Eulerian digraph G with n vertices and m arcs. If its minimum degree $\delta^+(G) \ge \lfloor \sqrt{m/n} \rfloor$, by Proposition 5.3.2 G already contains a cycle of length at least $1 + \lfloor \sqrt{m/n} \rfloor$. Therefore we can assume that there exists a vertex v with $\lfloor \sqrt{m/n} \rfloor > d^+(v) := t$. As G is Eulerian, there exist t arc-disjoint cycles $C_1, C_2, ..., C_t$ passing through v. If one of these cycles has length at least $\lfloor \sqrt{m/n} \rfloor + 1$ then again we are done. Otherwise, $|C_i| \le \lfloor \sqrt{m/n} \rfloor$ for all $1 \le i \le t$. Now we delete from G the vertex v together with the arcs of the cycles C_1, \cdots, C_t . The resulting Eulerian digraph has n-1 vertices and m' arcs, where $m' = m - \sum_{i=1}^t |C_i| \ge m - t \lfloor \sqrt{m/n} \rfloor \ge m(1-\frac{1}{n})$. By the inductive hypothesis, the new digraph (therefore G) has a cycle of length at least

$$1 + \sqrt{m'/(n-1)} \ge 1 + \sqrt{m(1-\frac{1}{n})/(n-1)} \ge 1 + \lfloor \sqrt{m/n} \rfloor.$$

5.4 Concluding remarks

We end with some remarks on the Bollobás-Scott conjecture whose unweighted version states that an Eulerian digraph with n vertices and m arcs has a cycle of length $\Omega(m/n)$. The "canonical" proof for showing that an undirected graph with this many vertices and edges has a cycle of length m/n proceeds by first passing to a subgraph G' with minimum degree at least m/n and then applying Proposition 5.3.2 to G'. We can then interpret the second statement of Theorem 5.1.4 as stating that when applied to Eulerian digraphs, this approach can only produce cycles of length $O(m^2/n^3)$.

There is, however, another way to show that an undirected graph has a cycle of length m/n using DFS. Recall that the DFS (Depth First Search) is a graph algorithm that visits all the vertices of a (directed or undirected) graph G as follows. It maintains three sets of vertices, letting S be the set of vertices which we have completed exploring them, T be the set of unvisited vertices, and $U = V(G) \setminus (S \cup T)$, where the vertices of U are kept in a *stack* (a *last in, first out* data structure). The DFS starts with $S = U = \emptyset$ and T = V(G).

While there is a vertex in $V(G) \setminus S$, if U is non-empty, let v be the last vertex that was added to U. If v has a neighbor $u \in T$, the algorithm inserts u to U and repeats this step. If v does not have a neighbor in T then v is popped out from U and is inserted to S. If U is empty, the algorithm chooses an arbitrary vertex from T and pushes it to U. Observe crucially that all the vertices in U form a directed path, and that there are no edges from S to T.

Consider any DFS tree T rooted at some vertex v. Recall that any edge of G is either an edge of T or a backward edge, that is, an edge connecting a vertex v to one of its ancestors in T. Hence, if G has no cycle of length at least t, then any vertex of T sends at most t-1 edges to his ancestors in T. This means that $m \leq nt$ or that $t \geq m/n$. Note that this argument shows that any DFS tree of an undirected graph has depth at least m/n. It is

thus natural to try and adapt this idea to the case of Eulerian digraphs. Unfortunately, as the following proposition shows, this approach fails in Eulerian digraphs.

Proposition 5.4.1. There is an Eulerian digraph G with average degree at least $\sqrt{n}/20$ such that some DFS tree of G has depth 4.

Proof. We first define a graph G' as follows. Let t be a positive integer and let G' be a graph consisting of 2t vertex sets V_1, \ldots, V_{2t} , each of size t. We also have a special vertex r, so G' has $2t^2 + 1$ vertices. We now define the arcs of G' using the following iterative process. We have t iterations, where in iteration $1 \leq j \leq t$ we add the following arcs; we have t arcs pointing from r to the t vertices of V_j , then a matching between the t vertices of V_j to the vertices of V_{j+1} , and in general a matching between V_k to V_{k+1} for every $j \leq k \leq 2t - j$. We finally have t arcs from V_{2t-j+1} to r. We note that we can indeed add a new (disjoint from previous ones) matching between any pair of sets (V_k, V_{k+1}) in each of the t iterations by relying on the fact that the edges of the complete bipartite graph $K_{t,t}$ can be split into tperfect matchings. Observe that in iteration j we add t(2t - 2j + 3) arcs to G'. Hence G'has

$$\sum_{j=1}^{t} t(2t - 2j + 3) \ge t^3$$

arcs. Moreover it is easy to see from construction that G' is Eulerian. To get the graph G we modify G' as follows; for every vertex $v \in \bigcup_{i=1}^{2t} V_i$ we add two new vertices v^{in}, v^{out} and add a 4-cycle $(r, v^{in}, v, v^{out}, r)$. We get that G has $6t^2 + 1$ vertices and more than t^3 arcs, so setting $n = 6t^2 + 1$ we see that G has average degree at least $\sqrt{n}/20$.

Now consider a DFS tree of G which proceeds as follows; we start at r, and then for every $v \in V_{2t}$ go to v^{in} then to v and then to v^{out} . Next, for every $v \in V_{2t-1}$ we go to v^{in} then to v and then to v^{out} . We continue this way until we cover all the vertices of G. The DFS tree we thus get has r as its root, and $2t^2$ paths of length 3 (of type r, v^{in}, v, v^{out}) attached to it. \Box

Observe that the above proposition does not rule out the possibility that some DFS tree has depth $\Omega(m/n)$. We note that proving such a claim will imply that an Eulerian digraph has a path of length $\Omega(m/n)$. It appears that even this special case of the Bollobás-Scott conjecture is still open, so it might be interesting to further investigate this problem. In fact, we suspect that if G is a connected Eulerian digraph then for any vertex $v \in G$ there is a path of length $\Omega(m/n)$ starting at v. This statement for undirected graphs follows from the DFS argument at the beginning of this section.

Acknowledgment The author would like to thank Jacques Verstraete for helpful initial discussions.

CHAPTER 6

Quasi-randomness of graph balanced cut properties

6.1 Introduction

The study of random structures has seen a tremendous success in modern combinatorics and theoretical computer science. One example is the Erdős-Rényi random graph G(n, p)proposed in the 1950's and intensively studied thereafter. G(n, p) is the probability space of graphs over n vertices where each pair of vertices forms an edge independently with probability p. Random graphs are not only an interesting object of study on their own but also proved to be a powerful tool in solving numerous open problems. The success of random structures served as a natural motivation for the following question: How can one tell when a given structure behaves like a random one? Such structures are called *quasi-random*. In this chapter, we study quasi-random graphs, which, following Thomason [98, 99], can be informally defined as graphs whose edge distribution closely resembles that of a random graph (the formal definition will be given later). One fundamental result in the study of quasi-random graphs is the following theorem proved by Chung, Graham and Wilson [19].

Theorem 6.1.1. Fix a real $p \in (0,1)$. For an n-vertex graph G, define e(U) to be the number of edges in the induced subgraph spanned by vertex set U, then the following properties are equivalent.

 \mathcal{P}_1 : For any subset of vertices $U \subset V(G)$, we have $e(U) = \frac{1}{2}p|U|^2 \pm o(n^2)$. $\mathcal{P}_2(\alpha)$: For any subset of vertices $U \subset V(G)$ of size αn , we have $e(U) = \frac{1}{2}p|U|^2 \pm o(n^2)$.

 \mathcal{P}_3 : $e(G) = \frac{1}{2}pn^2 \pm o(n^2)$ and G has $\frac{1}{8}p^4n^4 \pm o(n^4)$ cycles of length 4.

Throughout this chapter, unless specified otherwise, when considering a subset of vertices

 $U \subset V$ such that $|U| = \alpha n$ for some α , we tacitly assume that $|U| = \lfloor \alpha n \rfloor$ or $|U| = \lceil \alpha n \rceil$. Since we mostly consider asymptotic values, this difference will not affect our calculation.

For a positive real δ , we say that a graph G is δ -close to satisfying \mathcal{P}_1 if $e(U) = \frac{1}{2}p|U|^2 \pm \delta n^2$ for all $U \subset V(G)$, and similarly define it for other properties. The formal definition of equivalence of properties in Theorem 6.1.1 is as following: for every $\varepsilon > 0$, there exists a δ such that if a graph is δ -close to satisfying one property, then it is ε -close to satisfying another.

We call a graph *p*-quasi-random, or quasi-random if the density *p* is clear from the context, if it satisfies \mathcal{P}_1 , and consequently satisfies all of the equivalent properties of Theorem 6.1.1. We also say that a graph property is quasi-random if it is equivalent to \mathcal{P}_1 . Note that the random graph G(n, p) with high probability is *p*-quasi-random. However, it is not true that all the properties of random graphs are quasi-random. For example, it is easy to check that the property of having $\frac{1}{2}pn^2 + o(n^2)$ edges is not quasi-random (as an instance, there can be many isolated vertices). For more details on quasi-random graphs we refer the reader to the survey of Krivelevich and Sudakov [56]. Quasi-randomness was also studied in many other settings besides graphs, such as set systems [20], tournaments [21] and hypergraphs [22].

The main objective of this chapter is to study the quasi-randomness of graph properties given by certain graph cuts. These kind of properties were first studied by Chung and Graham in [20, 23]. For a real $\alpha \in (0, 1)$, the cut property $\mathcal{P}_C(\alpha)$ is the collection of graphs G satisfying the following: for any $U \subset V(G)$ of size $|U| = \alpha n$, we have $e(U, V \setminus U) = p\alpha(1 - \alpha)n^2 + o(n^2)$. As it turns out, for most values of α , the cut property $\mathcal{P}_C(\alpha)$ is quasi-random. In [20, 23], the authors proved the following beautiful theorem which characterizes the quasi-random cut properties.

Theorem 6.1.2. $\mathcal{P}_C(\alpha)$ is quasi-random if and only if $\alpha \neq 1/2$.

A cut is a partition of a vertex set V into subsets V_1, \dots, V_r , and if for a vector $\vec{\alpha} = (\alpha_1, \dots, \alpha_r)$, the size of the sets satisfies $|V_i| = \alpha_i |V|$ for all i, then we call this an $\vec{\alpha}$ -cut. An $\vec{\alpha}$ -cut is called *balanced* if $\vec{\alpha} = (1/r, \dots, 1/r)$ for some r, and is *unbalanced* otherwise. For a k-uniform hypergraph G and a cut V_1, \dots, V_r of its vertex set, let $e(V_1, \dots, V_r)$

be the number of hyperedges which have at most one vertex in each part V_i for all i.

A k-uniform hypergraph G is p-quasi-random if for every subset of vertices $U \subset V(G)$, $e(U) = p \frac{|U|^k}{k!} \pm o(n^k)$. Let $\mathcal{P}_C(\vec{\alpha})$ be the following property: for every $\vec{\alpha}$ -cut V_1, \ldots, V_r , $e(V_1, \cdots, V_r) = (p + o(1))n^k \sum_{S \subset [r], |S| = k} \prod_{i \in S} \alpha_i$. Shapira and Yuster [90] generalized Theorem 6.1.2 by proving the following theorem.

Theorem 6.1.3. Let $k \ge 2$ be a positive integer. For k-uniform hypergraphs, the cut property $\mathcal{P}_C(\vec{\alpha})$ is quasi-random if and only if $\vec{\alpha} \ne (1/r, \ldots, 1/r)$ for some $r \ge k$.

For a fixed graph H, let \mathcal{P}_H be the following property : for every subset $U \subset V$, the number of copies of H in U is $(p^{|E(H)|} + o(1)) \binom{|U|}{|V(H)|}$. In [94], Simonovits and Sós proved that \mathcal{P}_H is equivalent to \mathcal{P}_1 and hence is quasi-random. For a fixed graph H, as a common generalization of Chung and Graham's and Simonovits and Sós' theorems, we can consider the number of copies H having one vertex in each part of a cut. Let us consider the cases when H is a clique of size k.

Definition 6.1.4. Let k, r be positive integers such that $r \ge k \ge 2$, and let $\vec{\alpha} = (\alpha_1, \dots, \alpha_r)$ be a vector of positive real numbers satisfying $\sum_{i=1}^r \alpha_i = 1$. We say that a graph satisfies the K_k cut property $C_k(\vec{\alpha})$ if for every $\vec{\alpha}$ -cut (V_1, \dots, V_r) , the number of copies of K_k which have at most one vertex in each of the sets V_i is $(p^{\binom{k}{2}} \pm o(1))n^k \sum_{S \subset [r], |S|=k} \prod_{i \in S} \alpha_i$.

Shapira and Yuster [90] proved that for $k \geq 3$, $C_k(\vec{\alpha})$ is quasi-random if $\vec{\alpha}$ is unbalanced (note that $C_2(\vec{\alpha})$ is quasi-random if and only if $\vec{\alpha}$ is unbalanced). This result is a corollary of Theorem 6.1.3 by the following argument. For a graph G satisfying $C_k(\vec{\alpha})$, consider the kuniform hypergraph G' on the same vertex set where a k-tuple of vertices forms an hyperedge if and only if they form a clique in G. Then G' satisfies $\mathcal{P}_C(\vec{\alpha})$ and thus is quasi-random. By the definition of the quasi-randomness of hypergraphs, this in turn implies that the number of cliques of size k inside every subset of V(G) is "correct", and thus by Simonovits and Sós' result, G is quasi-random.

Note that for balanced $\vec{\alpha}$ this approach does not give enough information, since it is not clear if there exists a graph whose hypergraph constructed by the above mentioned process is not quasi-random but satisfies $\mathcal{P}_C(\vec{\alpha})$ (nonetheless as the reader might suspect, the properties $\mathcal{P}_C(\vec{\alpha})$ and $\mathcal{C}_k(\vec{\alpha})$ are closely related even for balanced $\vec{\alpha}$). Shapira and Yuster made this observation and left the balanced case as an open question asking whether it is quasi-random or not. Janson [48] independently posed the same question in his paper that studied quasi-randomness under the framework of graph limits. In this chapter, we settle this question by proving the following theorem:

Theorem 6.1.5. Fix a real $p \in (0,1)$ and positive integers r, k such that $r \ge k \ge 3$. For every $\varepsilon > 0$, there exists a positive real δ such that the following is true. If G is a graph which has density p and is δ -close to satisfying the K_k balanced cut property $C_k(1/r, \dots, 1/r)$, then G is ε -close to being p-quasi-random.

The rest of the chapter is organized as follows. In Section 6.2 we introduce the notations we are going to use throughout the chapter and state previously known results that we need later. In Section 6.3 we give a detailed proof of the most important base case of Theorem 6.1.5, triangle balanced cut property, i.e. $C_3(1/r, \dots, 1/r)$. In Section 6.4, we prove the general case as a consequence of the base case. The last section contains some concluding remarks and open problems for further study.

6.2 Preliminaries

Given a graph G = (V, E) and two vertex sets $X, Y \subset V(G)$, we denote by E(X, Y)the set of edges which have one end point in X and the other in Y. Also we write e(X,Y) = |E(X,Y)| to indicate the number of edges and $d(X,Y) = \frac{e(X,Y)}{|X||Y|}$ for the density. For a cut $\mathbf{X} = (X_1, \dots, X_r)$ of the vertex set, we say that a triangle with vertices u, v, w crosses the cut \mathbf{X} if it contains at most one vertex from each set, and denote it by $(u, v, w) \Leftrightarrow \mathbf{X}$. We use $Tr(\mathbf{X})$ for the number of triangles with vertices $(u, v, w) \Leftrightarrow \mathbf{X}$. For a k-uniform hypergraph and a partition V_1, \dots, V_t of its vertex set V into t parts, we define its density vector as the vector in $\mathbb{R}^{\binom{t}{k}}$ indexed by the k-subsets of [t] whose $\{i_1, \dots, i_k\}$ -entry is the density of hyperedges which have exactly one vertex in each of the sets V_{i_1}, \dots, V_{i_k} . Throughout the chapter, we always use subscripts such as $\delta_{6.2.6}$ to indicate that the parameter δ comes from Theorem 6.2.6.

To state asymptotic results, we utilize the following standard notations. For two positive-valued functions f(n) and g(n), write $f(n) = \Omega(g(n))$ if there exists a positive constant c such that $\liminf_{n\to\infty} f(n)/g(n) \ge c$, f(n) = o(g(n)) if $\limsup_{n\to\infty} f(n)/g(n) = 0$. Also, f(n) = O(g(n)) if there exists a positive constant C > 0such that $\limsup_{n\to\infty} f(n)/g(n) \le C$.

To isolate the unnecessary complication arising from the error terms, we will use the notation $x =_{\varepsilon} y$ if $|x - y| = O(\varepsilon)$ and say that x, y are ε -equal. For two vectors, we define $\vec{x} =_{\varepsilon} \vec{y}$ if $||\vec{x} - \vec{y}||_{\infty} = O(\varepsilon)$. We omit the proof of the following simple properties (we implicitly assume that the following operations are performed a constant number of times in total). Let C and c be positive constants.

(1a) (Finite transitivity) If $x =_{\varepsilon} y$ and $y =_{\varepsilon} z$, then $x =_{\varepsilon} z$.

(1b) (Complete transitivity) For a finite set of numbers $\{x_1, \dots, x_n\}$. If $x_i =_{\varepsilon} x_j$ for every

i, j, then there exists x such that $x_i =_{\varepsilon} x$ for all i.

(2) (Additivity) If $x =_{\varepsilon} z$ and $y =_{\varepsilon} w$, then $x + y =_{\varepsilon} z + w$.

(3) (Scalar product) If $x =_{\varepsilon} y$ and $0 < c \le a \le C$, then $ax =_{\varepsilon} ay$ and $x/a =_{\varepsilon} y/a$.

(4) (Product) If x, y, z, w are bounded above by C, then $x =_{\varepsilon} y$ and $z =_{\varepsilon} w$ implies that $xz =_{\varepsilon} yw$.

(5) (Square root) If both x and y are greater than c, then $x^2 =_{\varepsilon} y^2$ implies that $x =_{\varepsilon} y$.

(6) For the linear equation $A\vec{x} =_{\varepsilon} \vec{y}$, if all the entries of an invertible matrix A are bounded by C, and the determinant of A is bounded from below by c, then $\vec{x} =_{\varepsilon} A^{-1}\vec{y}$.

(7) If $xy =_{\varepsilon} 0$, then either $x =_{\sqrt{\varepsilon}} 0$ or $y =_{\sqrt{\varepsilon}} 0$.

6.2.1 Extremal Graph Theory

To prove the main theorem, we use the regularity lemma developed by Szemerèdi [97]. Let G = (V, E) be a graph and $\varepsilon > 0$ be fixed. A disjoint pair of sets $X, Y \subset V$ is called an ε -regular pair if $\forall A \subset X, B \subset Y$ such that $|A| \ge \varepsilon |X|, |B| \ge \varepsilon |X|$ satisfies $|d(X,Y) - d(A,B)| \leq \varepsilon$. A vertex partition $\{V_i\}_{i=1}^t$ is called an ε -regular partition if (i) V_i have equal size for all i, and (ii) (V_i, V_j) is ε -regular for all but at most εt^2 pairs $1 \leq i < j \leq n$. The regularity lemma states that every large enough graph admits a regular partition. In our proof, we use a slightly different form which can be found in [55]:

Theorem 6.2.1 (Regularity Lemma). For every real $\varepsilon > 0$ and positive integers m, r there exists constants $T(\varepsilon, m)$ and $N(\varepsilon, m)$ such that given any $n \ge N(\varepsilon, m)$, the vertex set of any n-vertex graph G can be partitioned into t sets V_1, \dots, V_t for some t divisible by r and satisfying $m \le t \le T(\varepsilon, m)$, so that

- $|V_i| < [\varepsilon n]$ for every *i*.
- $||V_i| |V_j|| \le 1$ for all i, j.
- Construct a reduced graph H on t vertices such that $i \sim j$ in H if and only if (V_i, V_j) is ε -regular in G. Then the reduced graph has minimum degree at least $(1 - \varepsilon)t$.

As one can see in the following lemma, regular pairs are useful in counting small subgraphs of a graph (this lemma can easily be generalized to other subgraphs).

Lemma 6.2.2. Let V_1, V_2, V_3 be subsets of vertices. If the pair (V_i, V_j) is ε -regular with density d_{ij} for every distinct i, j, then the number of triangles $Tr(V_1, V_2, V_3)$ is

$$Tr(V_1, V_2, V_3) = (d_{12}d_{23}d_{31} + O(\varepsilon))|V_1||V_2||V_3|.$$

Proof. If a vertex $v \in V_1$ has degree $(1 + O(\varepsilon))d_{12}|V_2|$ in V_2 and $(1 + O(\varepsilon))d_{13}|V_3|$ in V_3 , then by the regularity of the pair (V_2, V_3) , there will be $(1 + O(\varepsilon))|V_2||V_3|d_{12}d_{23}d_{31}$ triangles which contain the vertex v. By the regularity of the pairs (V_1, V_2) and (V_1, V_3) , there are at least $(1 - 2\varepsilon)|V_1|$ such vertices. Moreover, there are at most $2\varepsilon|V_1||V_2||V_3|$ triangles which does not contain such vertex from V_1 . Therefore we have,

$$Tr(V_1, V_2, V_3) = (1 + O(\varepsilon))|V_1||V_2||V_3|d_{12}d_{23}d_{31} + 2\varepsilon|V_1||V_2||V_3| = (d_{12}d_{23}d_{31} + O(\varepsilon))|V_1||V_2||V_3|.$$

For a fixed graph H, a perfect H-factor of a large graph G is a collection of vertex disjoint copies of H that cover all the vertices of G. The next theorem is a classical theorem proved by Hajnal and Szemerédi [42] which establishes a sufficient minimum degree condition for the existence of a perfect clique factor.

Theorem 6.2.3. Let k be a fixed positive integer and n be divisible by k. If G is a graph on n vertices with minimum degree at least (1 - 1/k)n, then G contains a perfect K_k -factor.

6.2.2 Concentration

The following concentration result of Hoeffding [46] and Azuma [7] will be used several times during the proof (see also [72, Theorem 3.10]).

Theorem 6.2.4 (Hoeffding-Azuma Inequality). Let c_1, \ldots, c_n be constants, and let X_1, \ldots, X_n be a martingale difference sequence with $|X_k| \leq c_k$ for each k. Then for any $t \geq 0$,

$$Pr\left(\left|\sum_{i=1}^{n} X_{i}\right| \ge t\right) \le 2\exp\left(-\frac{t^{2}}{2\sum_{i=1}^{n} c_{i}^{2}}\right)$$

The next lemma is a corollary of Hoeffding-Azuma's inequality.

Lemma 6.2.5. Let G = (V, E) be a graph with |V| = n and $|E| = d\binom{n}{2}$ for some real d. Let U be a random subset of V constructed by selecting every vertex independently with probability α . Then $e(U) = \alpha^2 d\binom{n}{2} + o(n^2)$ with probability at least $1 - e^{-O(n^{1/2})}$.

Proof. Arbitrarily label the vertices by $1, \ldots, n$ and consider the vertex exposure martingale. More precisely, let X_k be the number of edges within U incident to k among the vertices $1, \ldots, k - 1$ ($X_k = 0$ if $k \notin U$). Then $e(U) = X_1 + \cdots + X_n$ and $(X_1 + \cdots + X_k - \mathbb{E}[X_1 + \cdots + X_k])_k$ forms a martingale such that $|X_k - \mathbb{E}[X_k]| \leq n$ for all k. Thus by Hoeffding-Azuma's inequality (Theorem 6.2.4),

$$Pr(|e(U) - \mathbb{E}[e(U)]| \ge C) \le 2e^{-2C^2/n^3}$$

Since $\mathbb{E}[e(U)] = \alpha^2 d\binom{n}{2}$, by selecting $C = n^{7/4}$, we obtain $e(U) = \alpha^2 d\binom{n}{2} + o(n^2)$ with probability at least $1 - e^{-O(n^{1/2})}$.

6.2.3 Quasi-randomness of hypergraph cut properties

Recall the cut property $\mathcal{P}_{C}(\vec{\alpha})$ defined in the introduction, and the fact that it is closely related to the clique cut property $\mathcal{C}_{k}(\vec{\alpha})$. While proving Theorem 6.1.3, Shapira and Yuster also characterized the structure of hypergraphs which satisfies the balanced cut property $\mathcal{P}_{C}(1/r, \dots, 1/r)$. Let $p \in (0, 1)$ be fixed and t be an integer. In order to classify the k-uniform hypergraphs satisfying the balanced cut property, we first look at certain edge-weighted hypergraphs. Fix a set $I \subset [t]$ of size |I| = t/2, and consider the weighted hypergraph on the vertex set [t] such that the hyperedge e has density $2p|e \cap I|/k$ for all e. Let $\mathbf{u}_{t,p,I}$ be the vector representing this weighted hypergraph, and let $W_{t,p}$ be the affine subspace of $\mathbb{R}^{\binom{t}{k}}$ spanned by the vectors $\mathbf{u}_{t,p,I}$ for all possible sets I of size |I| = t/2. In [90], the authors proved that the structure of a (non-weighted) hypergraph which is δ -close to satisfying the balanced cut property $\mathcal{P}_{C}(1/r, \dots, 1/r)$ can be described by the vector space $W_{t,p}$ (note that the vector which has constant weight lies in this space).

Theorem 6.2.6. Let $p \in (0,1)$ be fixed. There exists a real t_0 such that For every $\varepsilon > 0$, and for every $t \ge t_0$ divisible by $2r^{-1}$, there exists $\delta = \delta(t, \varepsilon) > 0$ so that the following holds. If G is a k-uniform hypergraph with density p which is δ -close to satisfying the balanced cut property $\mathcal{P}_C(1/r, \dots, 1/r)$, then for any partition of V(G) into t equal parts, the density vector **d** of this partition satisfies $\|\mathbf{d} - \mathbf{y}\|_{\infty} \le \varepsilon$ for some vector $\mathbf{y} \in W_{t,p}$.

A part of the proof of Shapira and Yuster's theorem relies on showing that certain matrices have full rank, and they establish this result by using the following famous result from algebraic combinatorics proved by Gottlieb [40]. For a finite set T and integers h and ksatisfying $|T| > h \ge k \ge 2$, denote by B(T, h, k) the h versus k inclusion matrix of T which is the $\binom{|T|}{h} \times \binom{|T|}{k}$ 0-1 matrix whose rows are indexed by the h-element subsets of T, columns are indexed by the k-elements subsets of T, and entry (I, J) is 1 if and only if $J \subset I$.

Theorem 6.2.7. $rank(B(T, h, k)) = \binom{|T|}{k}$ for all $|T| \ge h + k$.

¹ The authors omitted the divisibility condition in their paper [90].

6.3 Base case - Triangle Balanced Cut

In this section we prove a special case, triangle balanced cut property, of the main theorem. Our proof consists of several steps. Let G be a graph which satisfies the triangle balanced cut property. First we apply the regularity lemma to describe the structure of G by an ε -regular partition $\{V_i\}_{i=1}^t$. This step allows us to count the edges or triangles effectively using regularity of the pairs. From this point on, we focus only on the cuts whose parts consist of a union of the sets V_i . In the next step, we swap some vertices of V_i and V_j . By the triangle cut property, we can obtain an algebraic relation of the densities inside V_i and between V_i and V_j . After doing this, the problem is transformed into solving a system of nonlinear equations, which basically implies that inside any clique of the reduced graph, most of the densities are very close to each other. Finally resorting to results from extremal graph theory, we can conclude that almost all the densities are equal and thus prove the quasi-randomness of triangle balanced cut property.

Theorem 6.3.1. Fix a real $p \in (0,1)$ and an integer $r \ge 3$. For every $\varepsilon > 0$, there exists a positive real δ such that the following is true. If G is a graph which has density p and is δ -close to satisfying the triangle balanced cut property $C_3(1/r, \dots, 1/r)$, then G is ε -close to being p-quasi-random.

Let G be a graph δ -close to satisfying $C_3(1/r, \dots, 1/r)$. By applying the regularity lemma, Theorem 6.2.1, to G, we get an ε -regular equipartition $\pi = \{V_i\}_{i=1}^t$. We can assume that $|V_1| = \dots = |V_t|$ by deleting at most t vertices. The reason this can be done is that later when we use the triangle cut property to count the number of triangles, the error term that this deletion creates is at most tn^2 which is negligible comparing to δn^3 when n is sufficiently large. Also in the definition of quasi-randomness, the error term from counting edges is at most tn, which is also $o(n^2)$.

Now denote the edge density within V_i by x_i , the edge density between V_i and V_j by d_{ij} , and the density of triangles in the tripartite graph formed by (V_i, V_j, V_k) by d_{ijk} . Call a triple (V_i, V_j, V_k) regular if each of the three pairs is regular. Consider a family $\{\pi_{\alpha}\}_{\alpha \in [0,1]}$ of partitions of G given as follows:

$$\pi_{\alpha} = ((1 - \alpha)V_1 + \alpha V_2, \alpha V_1 + (1 - \alpha)V_2, V_3, \cdots V_t).$$

In other words, we pick U_1 and U_2 both containing α -proportion of vertices in V_1 and V_2 uniformly at random and exchange them to form a new equipartition π_{α} . To be precise, for fixed α , the partition π_{α} is not a well-defined partition. However, for convenience we assume that π_{α} is a partition constructed as above which satisfies some explicit properties that we soon mention which a.a.s. holds for random partitions. Denote the new triangle density vector of π_{α} by $\mathbf{d}^{\alpha} = (d_{ijk}^{\alpha})$.

We know that every $(1/r, \dots, 1/r)$ -cut $\mathbf{X} = (X_1, \dots, X_r)$ of the vertex set [t] also gives a $(1/r, \dots, 1/r)$ -cut of V(G). By the triangle balanced cut property, for every $\alpha \in [0, 1]$,

$$(p^{3} \pm \delta) \left(\frac{n}{r}\right)^{3} \cdot {\binom{r}{3}} = \sum_{(i,j,k) \in \mathbf{X}} Tr(X_{i}, X_{j}, X_{k}) = \sum_{(i,j,k) \in \mathbf{X}} d_{ijk}^{\alpha} \left(\frac{n}{t}\right)^{3}.$$

So $\sum_{(i,j,k) \in \mathbf{X}} d_{ijk}^{\alpha} = (p^3 \pm \delta) {\binom{r}{3}} {\binom{t}{r}}^3$. Let M be the $\binom{t}{t/r, \cdots, t/r} \times {\binom{t}{3}} 0 - 1$ matrix whose rows are indexed by the $(1/r, \cdots, 1/r)$ -cuts of the vertex set [t] and columns are indexed by the triples ${\binom{[t]}{3}}$. Where the $(\mathbf{X}, (i, j, k))$ -entry of M is 1 if and only if $(i, j, k) \in \mathbf{X}$. The observation above implies $M\mathbf{d}^{\alpha} = (p^3 \pm \delta) {\binom{r}{3}} {\binom{t}{2}}^3 \cdot \mathbf{1}$ where $\mathbf{1}$ is the all-one vector. Thus if we let $\mathbf{d}' = \mathbf{d}^{1/2} - \frac{1}{2}\mathbf{d}^0 - \frac{1}{2}\mathbf{d}^1$, then $M \mathbf{d}' =_{\delta t^3} \mathbf{0}$. From this equation we hope to get useful information about the densities x_i and d_{ij} . With the help of the following lemma, we can compute the new densities d_{ijk}^{α} , and thus the modified density vector \mathbf{d}' , in terms of the densities x_i and d_{ij} .

Lemma 6.3.2. Let ε satisfy $2\varepsilon < d_{ij}$ for every i, j and assume that the graph G is large enough. Then for all $\alpha \in (\varepsilon, 1-\varepsilon)$, there exists a choice of sets U_1, U_2 such that the following holds.

(1)

$$d_{ijk}^{\alpha} = \begin{cases} d_{ijk} & \text{if } \{i, j, k\} \cap \{1, 2\} = \emptyset \\ (1 - \alpha)d_{1jk} + \alpha d_{2jk} + o(1) & \text{if } i = 1 \text{ and } 2 \notin \{j, k\} \\ \alpha d_{1jk} + (1 - \alpha)d_{2jk} + o(1) & \text{if } i = 2 \text{ and } 1 \notin \{j, k\} \\ \text{see } (2) & \text{if } i = 1 \text{ and } j = 2 \end{cases}$$

(2) If (V_1, V_2, V_k) is a regular triple, then

$$d_{12k}^{\alpha} = ((1-\alpha)^2 + \alpha^2)d_{12}d_{1k}d_{2k} + \alpha(1-\alpha)(x_1d_{1k}^2 + x_2d_{2k}^2) + O(\varepsilon).$$

(3) Let $d'_{ijk} = d^{\alpha}_{ijk} - (1 - \alpha)d^{0}_{ijk} - \alpha d^{1}_{ijk}$. Then

$$d'_{ijk} = \begin{cases} 0 & \text{if } \{i, j, k\} \cap \{1, 2\} = \emptyset \\ o(1) & \text{if } i = 1 \text{ and } 2 \notin \{j, k\} \\ o(1) & \text{if } i = 2 \text{ and } 1 \notin \{j, k\} \end{cases}$$

Moreover, for the case i = 1 and j = 2, if (V_1, V_2, V_k) is a regular triple, then

$$d'_{12k} = \alpha(1-\alpha)(x_1d_{1k}^2 + x_2d_{2k}^2 - 2d_{12}d_{1k}d_{2k}) + O(\varepsilon).$$

Proof. (1) The claim clearly holds for the cases $\alpha = 0$ and $\alpha = 1$.

For $\alpha \in (0, 1)$, if $\{i, j, k\} \cap \{1, 2\} = \emptyset$, then the density d_{ijk}^{α} is not affected by the swap of vertices in V_1 and V_2 so it remains the same with d_{ijk} . In the case that $\{i, j, k\} \cap \{1, 2\} = \{1\}$, without loss of generality we assume i = 1 and $j, k \neq 2$. We also assume that there are S_x triangles with a fixed vertex $x \in V_1 \cup V_2$ and two other vertices belonging to V_j and V_k respectively (note that $S_x \leq m^2$). After swapping subset $U_1 \subset V_1$ with $U_2 \subset V_2$ such that $|U_1| = |U_2| = \alpha |V_i|$, we know that the number of triangles in triple (($(V_1 \cup U_2) \setminus U_1$), V_j, V_k) changes by $\sum_{u \in U_2} S_u - \sum_{u \in U_1} S_u$.

Assume $|V_i| = m$, instead of taking αm vertices uniformly at random, take every vertex in V_1 (or V_2) independently with probability α . This gives random variables X_i for $1 \le i \le m$ having Bernoulli distribution with parameter α . Let $R = \sum_{i=1}^m X_i$ and $S = \sum_{i=1}^m X_i S_i$.

These random variables represent the number of vertices chosen for U_1 , and the number of triangles in the triple that contain these chosen vertices, respectively. It is easy to see

$$Pr(R = \alpha m) = \alpha^{\alpha m} (1 - \alpha)^{(1 - \alpha)m} \binom{m}{\alpha m} \sim \Omega\left(\frac{1}{\sqrt{\alpha(1 - \alpha)}} m^{-1/2}\right),$$

and by Hoeffding-Azuma's inequality (Theorem 6.2.4)

$$Pr(|S - \mathbb{E}S| \ge C) \le 2\exp\left(-\frac{C^2}{2\sum_{i=1}^m S_i^2}\right) \le 2\exp\left(-\frac{C^2}{2m^3}\right).$$

Let $C = m^2$, and the second probability decreases much faster than the first probability, thus we know that conditioned on the event $R = \alpha m$, S is also concentrated at its expectation $\mathbb{E}S = \sum_{i=1}^{m} \alpha S_x = \alpha d_{1jk} m^3$. From here we know the number of triangles changes by

$$\sum_{u \in U_2} S_u - \sum_{u \in U_1} S_u = \alpha d_{2jk} m^3 - \alpha d_{1jk} m^3 + o(m^3)$$

Therefore the new density is

$$d_{1jk}^{\alpha} = d_{1jk} + \left(\sum_{u \in U_2} S_u - \sum_{u \in U_1} S_u\right)/m^3 = (1 - \alpha)d_{1jk} + \alpha d_{2jk} + o(1)$$

We can use similar method to compute d^{α}_{2jk} .

(2) Let $U_1 \subset V_1$, $U_2 \subset V_2$ be as in (1), and let $V'_1 = (V_1 \setminus U_1) \cup U_2$, $V'_2 = (V_2 \setminus U_2) \cup U_1$. Then we have the identity

$$Tr(V_1', V_2', V_k) = Tr(U_1, U_2, V_k) + Tr(U_1, V_1 \setminus U_1, V_k) + Tr(V_2 \setminus U_2, U_2, V_k) + Tr(V_2 \setminus U_2, V_1 \setminus U_1, V_k) + Tr(V_2 \setminus U_2, V_2, V_1 \setminus U_1, V_k) + Tr(V_2 \setminus U_2, V_k) + Tr(V_2 \setminus U_k) + Tr(V_2 \setminus U_k) + Tr(V_k \setminus U_k) + Tr(V_$$

Since $\alpha \in (\varepsilon, 1 - \varepsilon)$, the triples (U_1, U_2, V_k) and $(V_1 \setminus U_1, V_2 \setminus U_2, V_k)$ are regular. Thus by Lemma 6.2.2,

$$Tr(U_1, U_2, V_k) = (d_{12}d_{1k}d_{2k} + O(\varepsilon))|U_1||U_2||V_k| = (d_{12}d_{1k}d_{2k} + O(\varepsilon))\alpha^2 m^3,$$

and

$$Tr(V_1 \setminus U_1, V_2 \setminus U_2, V_k) = (d_{12}d_{1k}d_{2k} + O(\varepsilon))(1-\alpha)^2 m^3.$$

To compute $Tr(U_1, V_1 \setminus U_1, V_k)$, let $E_1^k \subset E(V_1)$ be the collection of edges which have $(d_{1k} \pm \varepsilon)^2 m$ common neighbors in V_k . By the regularity of the pair (V_1, V_k) , there are at

most $2\varepsilon m$ vertices in V_1 which does not have $(d_{1k} \pm \varepsilon)m$ neighbors in V_k . If v is not such a vertex, then by $\varepsilon < d_{1k} - \varepsilon$ and regularity, there are at most $2\varepsilon m$ other vertices in V_1 which do not have $(d_{1k} \pm \varepsilon)^2 m$ common neighbors with v. Consequently there are at most $4\varepsilon m^2$ edges inside V_1 which does not have $(d_{1k} \pm \varepsilon)^2 m$ common neighbors inside V_k , thus $|E_1^k| \ge x_1 {m \choose 2} - 4\varepsilon m^2$. By Lemma 6.2.5 and the calculation from part (1) there exists a choice of U_1 of size αm such that,

$$|E_1^k(U_1, V_1 \setminus U_1)| = |E_1^k| - |E_1^k(U_1)| - |E_1^k(V_1 \setminus U_1)|$$

= $(1 - \alpha^2 - (1 - \alpha)^2 + o(1))|E_1^k| = \alpha(1 - \alpha)x_1m^2 + O(\varepsilon)m^2.$

for all k. Note that the number of triangles $Tr(U_1, V_1 \setminus U_1, V_k)$ can be computed by adding the number of triangles containing the edges E_1^k and then the number of triangles containing at most $O(\varepsilon)m^2$ of the "exceptional" edges. The latter can be crudely bounded by $O(\varepsilon)m^2 \cdot m \leq O(\varepsilon)m^3$. Since each edge in E_1^k is contained in $(d_{1k} \pm \varepsilon)^2 m$ triangles (within the triple (V_1, V_2, V_k)),

$$Tr(U_1, V_1 \setminus U_1, V_k) = |E_1^k(U_1, V_1 \setminus U_1)| \cdot (d_{1k}^2 + O(\varepsilon))m + O(\varepsilon)m^3$$

= $\alpha (1 - \alpha) x_1 d_{1k}^2 m^3 + O(\varepsilon)m^3.$

Similarly we can show that there exists a choice of U_2 of size αm that gives

$$Tr(U_2, V_2 \setminus U_2, V_k) = \alpha(1 - \alpha)x_2d_{2k}^2m^3 + O(\varepsilon)m^3$$

for all k. Combining all the results together, we can conclude the existence of sets U_1 , U_2 such that

$$d_{12k}^{\alpha} = ((1-\alpha)^2 + \alpha^2)d_{12}d_{1k}d_{2k} + \alpha(1-\alpha)(x_1d_{1k}^2 + x_2d_{2k}^2) + O(\varepsilon).$$

Part (3) is just a straightforward computation from the definition of d'_{ijk} and (2).

Lemma 6.3.3. Let ε satisfy $2\varepsilon < d_{ij}$ for every i, j and assume that the graph G is large enough. If (V_i, V_j, V_k) is a regular triple, then $x_i d_{ik}^2 + x_j d_{jk}^2 - 2d_{ij} d_{ik} d_{jk} =_{\delta t^3 + \varepsilon} 0$.

Proof. As mentioned before Lemma 6.3.2, the vector $\mathbf{d}' = \mathbf{d}^{1/2} - \frac{1}{2}\mathbf{d}^0 - \frac{1}{2}\mathbf{d}^1$ satisfies $M\mathbf{d}' =_{\delta t^3} \mathbf{0}$. For an index $k \neq 1, 2$, consider a balanced partition \mathbf{X} of the vertex set [t]
such that 1 and 2 lies in different parts, and let Y be the union of the parts which contains neither 1 nor 2. Then by Lemma 6.3.2 (3),

$$0 =_{\delta t^3} \sum_{(i,j,k) \models \mathbf{X}} d'_{ijk} = t^3 \cdot o(1) + \sum_{k \in Y} d'_{12k},$$

where o(1) goes to 0 as the number of vertices in the graph G grows. Since Y can be an arbitrary set of size (r-2)t/r not containing 1 and 2, this immediately implies that $d'_{12k} =_{\delta t^3} 0$ for all k. Thus if (V_1, V_2, V_k) is a regular triple, then by Lemma 6.3.2 (3), $x_1d^2_{1k} + x_2d^2_{2k} - 2d_{12}d_{1k}d_{2k} =_{\delta t^3+\varepsilon} 0$. By symmetry, we can replace 1 and 2 by arbitrary indices i, j.

Using Theorem 6.2.6 which characterizes the non-quasi-random hypergraphs satisfying the balanced cut property, we can prove the following lemma which allows us to bound the densities from below.

Lemma 6.3.4. There exists t_0 such that for fixed $p \in (0,1)$ and every $t \ge t_0$ which is divisible by 2r, there exist c = c(p) and $\delta_0 = \delta_0(t, p) > 0$ so that the following holds for every $\delta \le \delta_0$. If G is a graph with density p which is δ -close to satisfying the triangle balanced cut property, then for any partition π of V(G) into t equal parts, the density vector $\mathbf{d} = (d_{ij})_{i,j}$ satisfies $d_{ij} \ge c$ for all distinct $i, j \in [t]$.

Proof. Let $t_0 = t_{6.2.6}$, $\eta = p^3/6$, and for a given $t \ge t_0$ divisibly by 2r, let $\delta_0 = \min\{\delta_{6.2.6}(t,\eta), p^3/10\}$. Let V = V(G), and let G' be the hypergraph over the vertex set V such that $\{i, j, k\} \in E(G')$ if and only if i, j, k forms a triangle in the graph G. Let π be an arbitrary partition of V into t equal parts V_1, \ldots, V_t , and let $(d_{ij})_{i,j}$ be the density vector of the graph G, and $(d_{ijk})_{i,j,k}$ be the density vector of the hypergraph G' with respect to π . It suffices to show the bound $d_{ij} \ge p^3/10$ for every distinct $i, j \in [t]$. For simplicity we will only verify it for d_{12} . Note that number of triangles which cross V_1, V_2, V_k is at most $e(V_1, V_2) \cdot |V_k| = (|V_1||V_2|d_{12}) \cdot |V_k|$, and thus $d_{12k} \le d_{12}$ for all $k \ge 3$. Consequently, we can obtain the following inequality which would be crucial in our argument:

$$\sum_{k=3}^{t} d_{12k} \le (t-2) \cdot d_{12}. \tag{6.3.1}$$

Since G is δ -close to satisfying the triangle balanced cut property, we know that the density q of triangles is at least $q \ge p^3 - \delta$. By Theorem 6.2.6, $(d_{ijk})_{i,j,k}$ is η -equal to some vector in $W_{t,q}$. Recall that the vectors in $W_{t,q}$ can be expressed as an affine combination of the vectors $\mathbf{u}_{t,q,I} = (u_{ijk}^I)_{i,j,k}$ for sets $I \subset [t]$ of size |I| = t/2, and note that the following is true no matter how we choose the set I:

$$\sum_{k=3}^{t} u_{12k}^{I} \ge \sum_{k \in I} \frac{2q}{3} \ge \left(\frac{t}{2} - 2\right) \frac{2q}{3}.$$

Since $(d_{ijk})_{i,j,k}$ is ε -equal to an affine combination of these vectors, for large enough t we have

$$\sum_{k=3}^{t} d_{12k} \ge \left(\frac{t}{2} - 2\right) \frac{2q}{3} - t\varepsilon \ge \frac{tq}{3} - \frac{4q}{3} - \frac{tq}{6} \ge \frac{tq}{8}.$$
(6.3.2)

By combining (6.3.1) and (6.3.2), we obtain $d_{12} \ge q/8 \ge (p^3 - \delta)/8 \ge p^3/10$. Similarly we can deduce $d_{ij} \ge p^3/10$ for all distinct $i, j \in [t]$.

Since Lemma 6.3.4 asserts that all the pairwise densities d_{ij} are bounded from below by some constant, we are allowed to divide each side of an ε -equality by d_{ij} . This turns out to be a crucial ingredient in solving the equations given by Lemma 6.3.3.

Lemma 6.3.5. Given a positive real c and an integer $n \ge 4$, if for every distinct $i, j \in [n]$, $d_{ij} \ge c$ and for every distinct $i, j, k \in [n]$, $x_i d_{ik}^2 + x_j d_{jk}^2 - 2d_{ij} d_{ik} d_{jk} =_{\varepsilon} 0$, then there exists $s \in [n]$, x, y > 0 such that for any distinct $i, j \ne s$, we have $d_{ij} =_{\varepsilon} \sqrt{x}$. For any $i \ne s$, $d_{is} =_{\varepsilon} \sqrt{y}$. Moreover, $x_i =_{\varepsilon} \sqrt{x}$ if $i \ne s$ and $x_s =_{\varepsilon} \frac{\sqrt{x}}{y}(2y - x)$ (see, figure 6.1).

Proof. Throughout the proof, we heavily rely on properties of ε -equality given in Section 6.2.

First consider the case n = 4. By taking (i, j, k) = (1, 2, 3), (2, 3, 1), (3, 1, 2) respectively, we get the following system of equations:

$$\begin{cases} d_{13}^2 x_1 + d_{23}^2 x_2 =_{\varepsilon} 2d_{12}d_{13}d_{23} \\ d_{12}^2 x_1 + d_{23}^2 x_3 =_{\varepsilon} 2d_{12}d_{13}d_{23} \\ d_{12}^2 x_2 + d_{13}^2 x_3 =_{\varepsilon} 2d_{12}d_{13}d_{23} \end{cases}$$
(6.3.3)

Considering this as a system of linear equations with unknowns x_1, x_2, x_3 , the determinant of the coefficient matrix becomes $2d_{12}^2d_{13}^2d_{23}^2 \ge 2c^6$. Moreover, the coefficients in the matrix are bounded from above by 1. Therefore we can solve the linear system by appealing to property (6) of ε -equality and get

$$\begin{cases} x_1 =_{\varepsilon} \frac{d_{23}}{d_{12}d_{13}} (d_{12}^2 + d_{13}^2 - d_{23}^2) \\ x_2 =_{\varepsilon} \frac{d_{13}}{d_{12}d_{23}} (d_{12}^2 + d_{23}^2 - d_{13}^2) \\ x_3 =_{\varepsilon} \frac{d_{12}}{d_{13}d_{23}} (d_{13}^2 + d_{23}^2 - d_{12}^2) \end{cases}$$

$$(6.3.4)$$

Then

$$\begin{aligned} x_1 x_2 &=_{\varepsilon} \frac{d_{23}}{d_{12} d_{13}} (d_{12}^2 + d_{13}^2 - d_{23}^2) \cdot \frac{d_{13}}{d_{12} d_{23}} (d_{12}^2 + d_{23}^2 - d_{13}^2) \\ &=_{\varepsilon} \frac{1}{d_{12}^2} [d_{12}^4 - (d_{13}^2 - d_{23}^2)^2] \\ &=_{\varepsilon} \frac{1}{d_{12}^2} [d_{12}^4 - (d_{14}^2 - d_{24}^2)^2]. \end{aligned}$$
(6.3.5)

The last equation comes from repeating the same step for the system of equations for indices 1, 2, and 4. Equation (6.3.5) implies $d_{13}^2 - d_{23}^2 =_{\varepsilon} \pm (d_{14}^2 - d_{24}^2)$, and $d_{ik}^2 - d_{jk}^2 =_{\varepsilon} \pm (d_{il}^2 - d_{jl}^2)$ for all distinct i, j, k, l in general. Assume that there exists an assignment $\{i, j, k, l\} = \{1, 2, 3, 4\}$ such that $d_{ik}^2 - d_{jk}^2 =_{\varepsilon} -(d_{il}^2 - d_{jl}^2) \neq_{\varepsilon} 0$, (we call such case as a "flip"). Without loss of generality let $d_{13}^2 - d_{23}^2 =_{\varepsilon} -(d_{14}^2 - d_{24}^2)$. By equation (6.3.3), we know that $x_1d_{12}^2 + x_3d_{23}^2 =_{\varepsilon} 2d_{12}d_{13}d_{23} =_{\varepsilon} x_2d_{12}^2 + x_3d_{13}^2$, from which we get

$$d_{12}^2(x_1 - x_2) =_{\varepsilon} (d_{13}^2 - d_{23}^2)x_3.$$

Replace the index 3 by 4 and we get

$$d_{12}^2(x_1 - x_2) =_{\varepsilon} (d_{14}^2 - d_{24}^2)x_4.$$

By the assumption on a "flip", by subtracting the two equalities we get $x_3 + x_4 =_{\varepsilon} 0$, thus $x_3 =_{\varepsilon} 0$ and $x_4 =_{\varepsilon} 0$. This is impossible from the equation $x_3 d_{13}^2 + x_4 d_{14}^2 =_{\varepsilon} 2d_{13} d_{14} d_{34}$ and the fact $d_{ij} \geq c$. Therefore no flip exists and we have

$$d_{ik}^2 - d_{il}^2 =_{\varepsilon} d_{jk}^2 - d_{jl}^2 \qquad \forall \{i, j, k, l\} = \{1, 2, 3, 4\}.$$
(6.3.6)

Since $d_{ij} \ge c$, the sum of $d_{12}^2 + d_{13}^2 - d_{23}^2$, $d_{12}^2 + d_{23}^2 - d_{13}^2$ and $d_{13}^2 + d_{23}^2 - d_{12}^2$ is equal to $d_{12}^2 + d_{13}^2 + d_{23}^2 \ge 3c^2$. So at least one of the terms is greater than c^2 , without loss of generality we can assume $d_{12}^2 + d_{13}^2 - d_{23}^2 \ge c^2$. Recall that $x_1 =_{\varepsilon} \frac{d_{23}}{d_{13}d_{12}} (d_{12}^2 + d_{13}^2 - d_{23}^2)$. By equation (6.3.6), we also have $x_1 =_{\varepsilon} \frac{d_{24}}{d_{14}d_{12}} (d_{12}^2 + d_{14}^2 - d_{24}^2) =_{\varepsilon} \frac{d_{24}}{d_{14}d_{12}} (d_{12}^2 + d_{13}^2 - d_{23}^2)$. Therefore $d_{23} = d_{24}$

$$\frac{d_{23}}{d_{13}} =_{\varepsilon} \frac{d_{24}}{d_{14}}.$$

By appealing to the bound $d_{ij} \geq c$, we get $d_{23}d_{14} =_{\varepsilon} d_{24}d_{13}$ and $d_{23}^2d_{14}^2 =_{\varepsilon} d_{24}^2d_{13}^2$, which implies

$$(d_{13}^2 - d_{23}^2)(d_{13}^2 - d_{14}^2)$$

= $d_{13}^2(d_{13}^2 - d_{23}^2) - d_{13}^2d_{14}^2 + d_{23}^2d_{14}^2$
= $_{\varepsilon} d_{13}^2(d_{14}^2 - d_{24}^2) - d_{13}^2d_{14}^2 + d_{13}^2d_{24}^2 = 0$

So either $d_{13}^2 = \sqrt{\varepsilon} d_{14}^2$ or $d_{13}^2 = \sqrt{\varepsilon} d_{23}^2$. Thus at this point we may assume the existence of indices i, j, k satisfying $d_{ik}^2 = \sqrt{\varepsilon} d_{jk}^2$. Assume that $d_{13}^2 = \sqrt{\varepsilon} d_{14}^2$ as the other case can be handled identically.

So $d_{13}^2 = \sqrt{\varepsilon} x$, $d_{14}^2 = \sqrt{\varepsilon} x$ for some x and by equation (6.3.6) we have $d_{23}^2 = \sqrt{\varepsilon} y$, $d_{24}^2 = \sqrt{\varepsilon} y$ for some y. We let $d_{34}^2 = z$, and the equation $d_{14}^2 - d_{34}^2 = d_{12}^2 - d_{32}^2$ given by (6.3.6) translates to $d_{12}^2 = \varepsilon x + y - z$. Moreover, from equation (6.3.4) for indices $\{1, 3, 4\}$ and $\{1, 2, 4\}$ we know that

$$x_1 =_{\varepsilon} \frac{d_{34}}{d_{14}d_{13}} (d_{14}^2 + d_{13}^2 - d_{34}^2) =_{\varepsilon} \frac{d_{24}}{d_{14}d_{12}} (d_{14}^2 + d_{12}^2 - d_{24}^2)$$

If we plug all equalities for d_{ij} into this identity, we get

$$(2x-z)\frac{\sqrt{z}}{x} =_{\sqrt{\varepsilon}} (2x-z)\frac{\sqrt{y}}{\sqrt{x(x+y-z)}}$$

So either $\frac{\sqrt{z}}{x} =_{\varepsilon^{1/4}} \frac{\sqrt{y}}{\sqrt{x(x+y-z)}}$ or $z =_{\varepsilon^{1/4}} 2x$. In the first case, by solving this equation we get either $z =_{\varepsilon^{1/8}} x$ or $z =_{\varepsilon^{1/8}} y$ (before multiplying each side of the equation by its denominators, one must establish the fact that x + y - z is bounded away from 0. This can be done by first figuring out the equation $\sqrt{\frac{z}{xy}} =_{\varepsilon^{1/4}} \sqrt{\frac{1}{x+y-z}}$, and then realizing that the left hand side is bounded from above). Both of the above solutions gives us a graph as claimed (see figure 6.1, for the case $z =_{\varepsilon^{1/8}} x$).



Figure 6.1: The structure of solution for n = 4.

In the second case $z =_{\varepsilon^{1/4}} 2x$, we consider the equation

$$x_2 =_{\varepsilon} \frac{d_{34}}{d_{24}d_{23}} (d_{24}^2 + d_{23}^2 - d_{34}^2) =_{\varepsilon} \frac{d_{14}}{d_{24}d_{12}} (d_{24}^2 + d_{12}^2 - d_{14}^2)$$

to get

$$(2y-z)\frac{\sqrt{z}}{x} =_{\sqrt{\varepsilon}} (2y-z)\frac{\sqrt{y}}{\sqrt{x(x+y-z)}}.$$

By the previous analysis we may assume $z = \sqrt{\varepsilon} 2y$, which implies $x =_{\varepsilon^{1/4}} y$ and $d_{12}^2 =_{\varepsilon^{1/4}} 0$. This is impossible by the fact $d_{12} \ge c$.

For n = 5, suppose not all the edge densities are ε -equal to the same value. In this case, there must be four vertices such that not all the densities between them are equal. Without loss of generality, we can assume $d_{12} =_{\varepsilon} d_{13} =_{\varepsilon} d_{14} =_{\varepsilon} x$, $d_{23} =_{\varepsilon} d_{24} =_{\varepsilon} d_{34} =_{\varepsilon} y$ and $x \neq y$. Now let us consider the collections of vertices $\{v_1, v_2, v_3, v_5\}$, $\{v_1, v_2, v_4, v_5\}$, $\{v_1, v_3, v_4, v_5\}$. From the case n = 4, we know that $d_{15} =_{\varepsilon} x$, $d_{25} =_{\varepsilon} d_{35} =_{\varepsilon} d_{45} =_{\varepsilon} y$. We can generalize it to arbitrary $n \geq 5$.

Note that if in the regular partition, every pair of sets were regular, then Lemma 6.3.5 itself forces the graph to be quasi-random, as apart from one part (which is negligible), all the densities are equal. However, the regularity lemma inevitably produces a partition which contains some irregular pairs, and in the remaining of the proof of Theorem 6.3.1 we will show how to handle this subtlety. The main idea is that since there are only a small number

of irregular pairs, the reduced graph will contain many cliques, and thus that we can use Lemma 6.3.5 to study its structure.

From now on in the reduced graph, when a clique of size at least 4 is given, we will call the exceptional vertex s to be "bad" and all others to be "good" vertices. We also call the densities d_s and d_{is} for any $i \neq s$ "bad" and d_{ij} "good" for $i, j \neq s$. However as it will later turn out, most cliques of size 4 has x = y, and in this case we call every vertex and edge to be "good".

Now we can combine Lemmas 6.3.3, 6.3.4, and 6.3.5 above to prove the main theorem which says that triangle balanced cut property is quasi-random.

Proof of Theorem 6.3.1 (triangle case). Let $c = c_{6.3.4}(p)$. We may assume that $\varepsilon < \min\{c/2, 1/4\}$. Let $t_0 = t_{6.3.4}$ and $T = T_{6.2.1}(\varepsilon, t_0)$. Let $\delta = \min_{t_0 \le t \le T}\{\varepsilon/t^3, \delta_{6.3.4}(t, p)\}$.

Let G be a graph which is δ -close to satisfying $C_3(1/r, \dots, 1/r)$. Consider the ε -regular equipartition π of G: $V(G) = V_1 \cup \dots \cup V_t$ we mentioned before. This gives a reduced graph H on t vertices of minimum degree at least $(1 - \varepsilon)t$ (we may assume that t is divisible by 4r). Every edge ij corresponds to an ε -regular pair (V_i, V_j) . We mark on each edge of H a weight d_{ij} which is the density of edges in (V_i, V_j) , and also the density x_i inside V_i on the vertices. Parameters are chosen so that Lemma 6.3.3 and Lemma 6.3.4 holds. Moreover, by the fact $\delta t^3 \leq \varepsilon$, we have $x_i d_{ik}^2 + x_j d_{jk}^2 - 2d_{ij} d_{ik} d_{jk} =_{\varepsilon} 0$ for every regular triple (V_i, V_j, V_k) . Thus whenever there is a clique of size at least 4 in H, by Lemma 6.3.5 we know that all the densities are ε -equal to each other, except for at most one "bad" vertex. Since $\varepsilon < 1/4$ and 4|t, we can apply Hajnal-Szemerédi theorem (Theorem 6.2.3) to the reduced graph H and get an equitable partition of the vertices of H into vertex disjoint 4-cliques $C_1, \dots, C_{t/4}$.

For every 4-clique C_i , from Lemma 6.3.5 we know that there is at most one "bad" vertex. For two 4-cliques C_i and C_j , we can consider the bipartite graph $\mathcal{B}(C_i, C_j)$ between them which is induced from H. If $\mathcal{B}(C_i, C_j) = K_{4,4}$, then it contains a subgraph isomorphic to $K_{2,2}$ where all the vertices are "good" (two vertices are good in C_i and other two in C_j). If we apply the structural lemma, Lemma 6.3.5, to this new 4-clique (together with two edges coming from the two known cliques), we get that the "good" densities of C_i and C_j are ε -equal to each other.

Now consider the reduced graph H' whose vertices correspond to the 4-cliques C_i , and C_i and C_j are adjacent in H' if and only if there is a complete bipartite graph between them. It is easy to see that $\delta(H') \geq (1 - 4\varepsilon)|H'|$. Take any two vertex $u', v' \in V(H')$, since d(u') + d(v') > |H'| they have a common neighbor w', and thus by the discussion above, the "good" density in $C_{u'}$ or $C_{v'}$ are ε -equal to the "good" density in $C_{w'}$. So all the "good" densities are ε -equal to each other. Thus by the total transitivity of ε -equality (see, Section 6.2), all the "good" densities are ε -equal to p' for some p'.

We would like to show that $d_{ij} =_{\varepsilon} p'$ for all but at most $O(\varepsilon)t^2$ edges of the reduced graph H. We already verified this for "good" edges $\{i, j\}$ belonging to the cliques $C_1, \ldots, C_{t/4}$. If C_i is adjacent to C_j in H' then they actually form a clique of size 8 in H, and by Lemma 6.3.5 there is at most one "bad" vertex there. Consequently the total number of cliques that contain at least one "bad" vertex cannot exceed the independence number of H', which is at most $|H'| - \delta(H') \leq \varepsilon t$. Thus among the cliques $C_1, \ldots, C_{t/4}$ there are at most εt cliques which contain at least one "bad" vertex. Moreover, the density of an edge in H which is part of a $K_{4,4}$ connecting two "good" cliques C_i and C_j are ε -equal to p' again by Lemma 6.3.5. Among the remaining edges, all but εt^2 are such edges as otherwise $e(H) < {t \choose 2} - \varepsilon t^2$ which is a contradiction. Therefore all but at most $O(\varepsilon)t^2$ edges of H have density ε -equal to p'. This in turn implies that the density of G is equal to $p' + O(\varepsilon)$. On the other hand we know that the density is p, thus $p' =_{\varepsilon} p$.

Now by verifying that G satisfies $\mathcal{P}_2(1/2)$ (see Theorem 6.1.1), we will show that G is quasi-random. For an arbitrary subset $U \subset V(G)$ of size n/2, let us compute the number of edges in e(U) and estimate its difference with the number of edges of a subset of size n/2 in

$$\begin{aligned} \left| e(U) - \binom{n/2}{2} p \right| \\ &\leq \sum_{i=1}^{t} \left| e(U \cap V_i) - \binom{|U \cap V_i|}{2} p \right| + \sum_{i,j} |e(U \cap V_i, U \cap V_j) - |U \cap V_i| |U \cap V_j| p | \\ &\leq \sum_{i=1}^{t} |V_i|^2 + \left(\sum_{i,j} |U \cap V_i| |U \cap V_j| \right) |p' - p + O(\varepsilon)| \\ &\leq n^2/t + O(\varepsilon |U|^2/2) = O(\varepsilon n^2). \end{aligned}$$
(6.3.7)

In the last equation, we took sufficiently large t depending on ε and p. Therefore by the quasi-randomness of $\mathcal{P}_2(1/2)$, we can conclude that G is a quasi-random graph. \Box

6.4 General Cliques

G(n,p).

Throughout the section, k and r are fixed integers satisfying $r \ge k \ge 4$. Let a r-balanced cut be a $(1/r, \dots, 1/r)$ -cut. In this section, we will prove the remaining cases of the main theorem, quasi-randomness of general k-clique r-balanced cut properties.

Theorem 6.4.1. Fix a real $p \in (0,1)$ and positive integers r, k such that $r \ge k \ge 4$. For every $\varepsilon > 0$, there exists a positive real δ such that the following is true. If G is a graph which has density p and is δ -close to satisfying the K_k balanced cut property $C_k(1/r, \dots, 1/r)$, then G is ε -close to being p-quasi-random.

Let G be a graph which is δ -close to satisfying the k-clique r-balanced cut property. Apply the regularity lemma (Theorem 6.2.1) to this graph to obtain an ε -regular partition $(V_i)_{i=1}^t$ of the vertex set. For $i \in [t]$, let x_i be the density of the edges within V_i , and for distinct $i, j \in [t]$, let d_{ij} be the density of the pair (V_i, V_j) . For $k \ge 2$, a k-tuple $J = (i_1, \ldots, i_k)$ is a multiset of k-indices (not necessarily distinct). Let d_J be the density of k-cliques which have exactly one vertices in each of the V_{i_a} for $a = 1, \ldots, k$. A k-tuple J is called *regular* if (V_{i_a}, V_{i_b}) forms an ε -regular pair for all $a, b \in [k]$. For a k-tuple J and a cut $\mathbf{X} = (X_1, \ldots, X_r)$, we say that J crosses the cut \mathbf{X} if $|J \cap X_i| \le 1$ for all i, and denote it by $J \Leftrightarrow \mathbf{X}$.

The proof of the k-clique r-balanced cut case follows the same line of the proof of the

triangle case. First we develop two lemmas which correspond to Lemma 6.3.3 and Lemma 6.3.4.

Lemma 6.4.2. Let ε be small enough depending on the densities d_{ij} for all $i, j \in [t]$. There exists a function $f : \mathbb{R} \to \mathbb{R}$ such that the following holds. Let J be a regular k-tuple, $J' \subset J$ be such that |J'| = k - 2, and $\{j_1, j_2\} = J \setminus J'$. Then,

$$x_{j_1}\left(\prod_{a\in J'} d_{aj_1}\right)^2 \left(\prod_{a,b\in J',a< b} d_{ab}\right) + x_{j_2}\left(\prod_{a\in J'} d_{aj_2}\right)^2 \left(\prod_{a,b\in J',a< b} d_{ab}\right) - 2\left(\prod_{a,b\in J,a< b} d_{ab}\right) =_{\varepsilon+\delta\cdot f(t)} 0.$$

Proof. For the sake of clarity, without loss of generality we consider the case $\{1, 2\} \subset J$ and $j_1 = 1, j_2 = 2$. As in the triangle case, by considering the family of *t*-partitions

$$\pi_{\alpha} = ((1-\alpha)V_1 + \alpha V_2, \alpha V_1 + (1-\alpha)V_2, V_3, \dots, V_t),$$

and the density vector $(d_J^{\alpha})_{J \in \binom{[t]}{r}}$ which arise from these partitions, we can define $\mathbf{d}' = (d'_J)_{J \in \binom{[t]}{r}}$ as $d'_J = d_J^{1/2} - \frac{1}{2}d_J^0 - \frac{1}{2}d_J^1$. The same proof as in Lemma 6.3.2 gives us,

$$d'_{J} = \begin{cases} 0 & \text{if } J \cap \{1, 2\} = \emptyset \\ o(1) & \text{if } J \cap \{1, 2\} = \{1\} \\ o(1) & \text{if } J \cap \{1, 2\} = \{2\} \end{cases}$$
(6.4.1)

and if $\{1,2\} \subset J$ (let $J' = J \setminus \{1,2\}$)) and J is a regular r-tuple, we have

$$d'_{J} = \alpha(1-\alpha) \left(\sum_{i=1}^{2} x_{i} \left(\prod_{a \in J'} d_{ai} \right)^{2} \left(\prod_{a,b \in J',a < b} d_{ab} \right) - 2 \left(\prod_{a,b \in J,a < b} d_{ab} \right) \right) + O(\varepsilon).$$
(6.4.2)

If $\{1,2\} \subset J$ and J is not a regular k-tuple, then we do not have any control on d'_J .

Let M be the $\binom{t}{t/r,\ldots,t/r} \times \binom{t}{k}$ 0-1 matrix whose rows are indexed by r-balanced cuts of the vertex set [t] and columns are indexed by the k-tuples $\binom{[t]}{k}$. Where the $((X_1, X_2, \ldots, X_r), J)$ -entry of M is 1 if and only if $J \Leftrightarrow (X_1, X_2, \ldots, X_r)$. We know that $M\mathbf{d}' =_{\delta t^k} \mathbf{0}$.

Consider the submatrix N of M formed by the rows of partitions which have 1 and 2 in different parts, and columns of r-tuples which include both 1 and 2. Let \mathbf{d}'' be the projection of \mathbf{d}' onto the coordinates corresponding to the r-tuples which contain both 1 and 2. By equation (6.4.1) and $M\mathbf{d}' =_{\delta t^k} \mathbf{0}$, we can conclude that $N\mathbf{d}'' =_{\delta t^k} \mathbf{0}$ given that the graph is large enough. Thus if we can show that N has full rank, then this implies that $d'_J =_{\delta \cdot f(t)} 0$ for all $\{1,2\} \subset J$ (appeal to property (6) of ε -equality - the entries of N are bounded and the size of N depends on t).

Observe that the matrix N can be considered as a 0-1 matrix whose rows are indexed by the collection of subsets $\mathbf{Y} = (Y_1, \ldots, Y_{r-2})$ of the set $T = \{3, \ldots, t\}$ where each part has size t/r, and columns are indexed by the (k-2)-tuples $J' \in \binom{T}{k-2}$, where the entry (\mathbf{Y}, J') is 1 if and only if $J' \cap \mathbf{Y}$. By fixing a subset of T of size (r-2)t/r and considering all possible \mathbf{Y} arising within this set, one can see that the row-space of N generates the row-space of the (r-2)t/r verses k-2 inclusion matrix of T, which we know by Gottlieb's theorem, Theorem 6.2.7, has full rank. This implies that N has full rank as well. \Box

Even though the equation which we obtained in Lemma 6.4.2 looks a lot more complicated than the triangle case, as it turns out, it is possible to make a substitution of variables so that the equations above become exactly the same as the equations in the triangle case. For a regular (k-3)-tuple I, an index $j \notin I$, and distinct $j_1, j_2 \notin I$, define

$$d_{j_1 j_2}^I := d_{j_1 j_2} \left(\prod_{a \in I} d_{a j_1} \right)^{1/2} \left(\prod_{a \in I} d_{a j_2} \right)^{1/2} \left(\prod_{a, b \in I, a < b} d_{a b} \right)^{1/3} \quad \text{and} \\ x_j^I := x_j \left(\prod_{a \in I} d_{a j} \right) \left(\prod_{a, b \in I, a < b} d_{a b} \right)^{1/3}.$$

Claim 6.4.3. Let *J* be a regular *k*-tuple, $I \subset J$ be of size |I| = k - 3, $\{j_1, j_2, j_3\} = J \setminus I$, and $J' = I \cup \{j_3\}$. Then

$$d_{j_1j_2}^I d_{j_2j_3}^I d_{j_3j_1}^I = \prod_{a,b \in J, a < b} d_{ab}, \quad and \quad x_{j_1}^I (d_{j_1j_3}^I)^2 = x_{j_1} \left(\prod_{a \in J'} d_{aj_1}\right)^2 \left(\prod_{a,b \in J', a < b} d_{ab}\right).$$

Proof. The claim follows from a direct calculation.

In other words, Claim 6.4.3 transforms the computation of the density of K_r in the graph into the computation of the density of triangles in another graph. This observation will greatly simplify the equations obtained from Lemma 6.4.2. **Lemma 6.4.4.** Let ε be small enough depending on the densities d_{ij} for all $i, j \in [t]$. There exists a function $f : \mathbb{R} \to \mathbb{R}$ such that the following holds. Let J be a regular k-tuple and $I \subset J$ be of size |I| = k - 3. For $\{j_1, j_2, j_3\} = J \setminus I$, we have $x_{j_1}^I (d_{j_1 j_3}^I)^2 + x_{j_2}^I (d_{j_2 j_3}^I)^2 - 2d_{j_1 j_2}^I d_{j_3 j_1}^I =_{\varepsilon + \delta \cdot f(t)} 0$.

Proof. This is an immediate corollary of Lemma 6.4.2 and Claim 6.4.3. \Box

Next lemma corresponds to Lemma 6.3.4 and establishes a lower bound on the densities. We omit the proof which is a straightforward generalization of the proof of Lemma 6.3.4.

Lemma 6.4.5. There exists t_0 such that for fixed $p \in (0,1)$ and every $t \ge t_0$ divisible by 2r, there exist c = c(k, p) and $\delta_0 = \delta_0(t, p) > 0$ so that the following holds for every $\delta \le \delta_0$. If G is a graph with density p which is δ -close to satisfying the k-clique balanced cut property, then for any partition π of V(G) into t equal parts, the density vector $\mathbf{d} = (d_{ij})_{i,j}$ satisfies $d_{ij} \ge c$ for all distinct $i, j \in [t]$.

For every fixed regular (k - 3)-tuple I, the set of equations that Lemma 6.4.4 gives is exactly the same as the set of equations obtained from Lemma 6.3.3. Consequently, by using Lemma 6.4.5, we can solve these equations for every fixed I just as in the triangle case.

Thus as promised, we can reduce the case of general cliques to the case of triangles. However, this observation does not immediately imply that $d_{j_1j_2} =_{\varepsilon} p$ for most of the pairs j_1, j_2 , since the only straightforward conclusion that we can draw is that for every regular (k-3)-tuple I, there exists a constant p_I such that $d_{j_1j_2}^I =_{\varepsilon} p_I$ for most of the pairs j_1, j_2 . In order to prove the quasi-randomness of balanced cut properties, we will need some control on the relation between different p_I . Call a k-tuple J excellent if it is regular, and for every (k-3)-tuple $I \subset J$, we have $d_{j_1j_2}^I =_{\varepsilon} p_I$ for all distinct $j_1, j_2 \in J \setminus I$.

Lemma 6.4.6. Let J be an excellent k-tuple. Then the density of every pairs in J are ε -equal to each other.

Proof. For the sake of clarity, assume that J = (1, 2, ..., k). First, consider I = (4, ..., k).

Then by the assumption, we have $d_{13}^I =_{\varepsilon} d_{23}^I$, which by definition gives,

$$d_{13} \left(\prod_{a \in I} d_{a1}\right)^{1/2} \left(\prod_{a \in I} d_{a3}\right)^{1/2} \left(\prod_{a,b \in I, a < b} d_{ab}\right)^{1/3} =_{\varepsilon} d_{23} \left(\prod_{a \in I} d_{a2}\right)^{1/2} \left(\prod_{a \in I} d_{a3}\right)^{1/2} \left(\prod_{a,b \in I, a < b} d_{ab}\right)^{1/3}$$

After cancelation of the same terms, we can rewrite this as,

$$d_{13}\left(\prod_{a=4}^{k} d_{a1}\right)^{1/2} =_{\varepsilon} d_{23}\left(\prod_{a=4}^{k} d_{a2}\right)^{1/2} \quad \Leftrightarrow \quad d_{13}^{1/2}\left(\prod_{a=3}^{k} d_{a1}\right)^{1/2} =_{\varepsilon} d_{23}^{1/2}\left(\prod_{a=3}^{k} d_{a2}\right)^{1/2}.$$
(6.4.3)

We can replace 3 by 4 up to k and multiply each side of all these equations to obtain,

$$\prod_{i=3}^{k} \left(d_{1i}^{1/2} \left(\prod_{a=3}^{k} d_{a1} \right)^{1/2} \right) =_{\varepsilon} \prod_{i=3}^{k} \left(d_{2i}^{1/2} \left(\prod_{a=3}^{k} d_{a2} \right)^{1/2} \right),$$

which is equivalent to

$$\left(\prod_{i=3}^{k} d_{1i}\right)^{(k-1)/2} =_{\varepsilon} \left(\prod_{i=3}^{k} d_{2i}\right)^{(k-1)/2}.$$

If we plug this back into equation (6.4.3), we get $d_{13} =_{\varepsilon} d_{23}$. By repeating this process for other choice of indices, we can conclude that the density of every pairs are ε -equal to each other.

We now combine all these observations to show that $d_e =_{\varepsilon} p$ for most of the edges e of the reduced graph, which will in turn imply the quasi-randomness.

Proof of Theorem 6.4.1. Choose ε_0 small enough depending on the constant $c = c_{6.4.5}(p)$ so that the condition of Lemma 6.4.4 holds, and let f be the function from Lemma 6.4.4. Let $\varepsilon \leq \min\{\varepsilon_0, 1/4\}, t_0 = t_{6.4.5}, \text{ and let } T = T_{6.2.1}(\varepsilon, t_0).$ Let $\delta = \min_{t_0 \leq t \leq T}\{\varepsilon/f(t), \delta_{6.4.5}(t, p)\}.$

Let G be a graph which is δ -close to satisfying the k-clique r-balanced cut property. Apply the regularity lemma (Theorem 6.2.1) to this graph to obtain an ε -regular partition $\{V_i\}_{i=1}^t$ of the vertex set where t is divible by 2r. For distinct $i, j \in [t]$, let d_{ij} be the density of the pair (V_i, V_j) . Note that the parameters are chosen so that Lemma 6.4.4 and Lemma 6.4.5 holds.

For every regular (k-3)-tuple I, define a graph H_I as following. The vertex set of H_I is the collection of elements of $[t] \setminus I$ which form a regular (k-2)-tuple together with I. And $j_1, j_2 \in V(H_I)$ forms an edge if and only if the (k-1)-tuple $I \cup (j_1, j_2)$ is regular. Since each part of the regular partition forms a regular pair with at least $(1 - \varepsilon)t$ of the other parts, we know that the graph H_I has at least $(1 - k\varepsilon)t$ vertices and minimum degree at least $(1 - 2k\varepsilon)t$. Thus by Lemma 6.4.4, Lemma 6.4.5, Lemma 6.3.5 and the proof of Theorem 6.3.1, we know that there exists a p_I such that at least $(1 - O(\varepsilon))$ -proportion of the edges of H_I have density ε -equal to p_I .

Select k indices j_1, \ldots, j_k out of [t] independently and uniformly at random. With probability at least $1 - O(\varepsilon)$, the k-tuple is regular. Moreover, with probability at least $1 - O(\varepsilon)$, $d_{j_1j_2}^{(j_4,\ldots,j_k)} =_{\varepsilon} p_{(j_4,\ldots,j_k)}$ and the same is true for other choices of indices as well. Therefore by the union bound, the k-tuple (j_1,\ldots,j_k) is excellent with probability at least $1 - O(\varepsilon)$. Equivalently, the number of excellent k-tuples is at least $(1 - O(\varepsilon)) {t \choose k}$.

Call a pair of indices in [t] excellent if it is contained in at least $\frac{2}{3} \binom{t}{k-2}$ excellent k-tuples. Assume that there are ηt^2 non-excellent edges. Then the number of non-excellent k-tuples are at least

$$\eta t^2 \times \frac{1}{3} {t \choose k-2} / {k \choose 2} = \Omega(\eta) {t \choose k}.$$

Therefore, $\eta = O(\varepsilon)$ and there are at most $O(\varepsilon)t^2$ non-excellent edges. We claim that all the excellent edges are ε -equal to each other. Take two excellent edges e, f. Since each of these edges form an excellent k-tuple with more than $\frac{2}{3} {t \choose k-2}$ of the (k-2)-tuples, there exists an (k-2)-tuple which forms an excellent k-tuple with both of these edges. Thus by Lemma 6.4.6 applied to each of these k-tuples separately, we can conclude that $d_e =_{\varepsilon} d_f$.

Consequently, by the total transitivity of ε -equality (see, Section 6.2), we can conclude that $d_e =_{\varepsilon} p'$ for some p' for every excellent edge e. Then apply the same reasoning as in the triangle case to show that $p' =_{\varepsilon} p$ and $G \in \mathcal{P}_2(1/2)$. This proves the quasi-randomness of the graph G.

6.5 Concluding Remarks

In this chapter, we proved the quasi-randomness of k-clique balanced cut properties for $k \geq 3$ and thus answered an open problem raised by both Shapira-Yuster [90] and Janson [48]. The most important base case was k = 3 where we solved a system of equations given by Lemma 6.3.3. The existence of "bad" vertex in Lemma 6.3.5 complicated the proof of the main theorem. It is hard to believe that the case can be significantly simplified since even if we assume that all the pairs are regular in the regular partition, there is an assignment of variables x_i and d_{ij} which is not all constant but forms a solution of the system.

We conclude this chapter with an open problem for further study.

Question 6.5.1. Let k, r be positive integers satisfying $r \ge k \ge 3$. Let H be a nonempty graph on k vertices, and assume that every $(1/r, \dots, 1/r)$ -cut of a graph G has the "correct" number of copies of H such that every vertex of H is in a different part of the cut. Does this condition force G to be quasi-random?

Acknowledgement The author would like to thank Asaf Shapira for the kindness and advice they provided that greatly helped us in doing this work, and Svante Janson for valuable corrections.

References

- [1] R. Aharoni and D. Howard, Size conditions for the existence of rainbow matchings, *in preparation.*
- [2] N. Alon, Neighborly families of boxes and bipartite coverings, Algorithms and Combinatorics 14 (1997), 27–31.
- [3] N. Alon, Ranking tournaments, SIAM J. Discrete Math., 20 (2006), no. 1, 137–142.
- [4] N. Alon, I. Haviv, private communication.
- [5] N. Alon and P. Seymour, A counterexample to the rank-coloring conjecture, *Journal of Graph Theory* 13(4) (1989), 523–525.
- [6] N. Alon and J. Spencer, The Probabilistic Method, John Wiley Inc., New York (2008).
- [7] K. Azuma, Weighted sums of certain dependent random variables, *Tôkuku Math. J.* 19 (1967), 357–367.
- [8] L. Babai and P. Frankl, Linear algebra methods in combinatorics with applications to geometry and computer science, The University of Chicago, 1992.
- [9] Z. Baranyai, On the factorization of the complete uniform hypergraph, Colloq. Math. Soc. János Bolyai 10 (1975), 91–108.
- [10] A. Bhattacharya, On a conjecture of Manickam and Singhi, Discrete Math. 272 (2003), 259–261.
- [11] T. Bier, A distribution invariant for the association schemes and strongly regular graphs, *Linear algebra and its applications* 57 (1984), 230–252.
- [12] T. Bier and N. Manickam, The first distribution invariant of the Johnson scheme, SEAMS Bull. Math. 11 (1987), 61–68.
- [13] B. Bollobás, Modern graph theory, Graduate Texts in Mathematics, vol. 184, Springer, New York, 1998.
- [14] B. Bollobás, D. Daykin and P. Erdős, Sets of independent edges of a hypergraph, Quart. J. Math. Oxford Ser. (2), 27 (1976), no. 105, 25–32.
- [15] B. Bollobás and A. Scott, A proof of a conjecture of Bondy concerning paths in weighted digraphs, J. Combin. Theory Ser. B, 66 (1996), no. 2, 283–292.
- [16] L. Caccetta and R. Häggkvist, On minimal digraphs with given girth, in Proc. 9th Southeastern Conference on Combinatorics, Graph Theory, and Computing (Boca Raton 1978), Congress. Numer. XXI 181–187.
- [17] P. Charbit, S. Thomassé and A. Yeo, The minimum feedback arc set problem is NP-hard for tournaments, *Combin. Probab. Comput.*, 16 (2007), 1–4.

- [18] M. Chudnovsky, P. Seymour and B. Sullivan, Cycles in dense digraphs, Combinatorica, 28 (2008) 1–18.
- [19] F. Chung, R. Graham, and R. Wilson, Quasi-random graphs, Combinatorica 9(4) (1989), 345–362.
- [20] F. Chung, R. Graham, and R. Wilson, Quasi-random set systems, Journal of the AMS 4 (1991), 151–196.
- [21] F. Chung and R. Graham, Quasi-random tournaments, Journal of Graph Theory 15 (1991), 173–198.
- [22] F. Chung and R. Graham, Quasi-random hypergraphs, Random Structures and Algorithms 1 (1990), 105–124.
- [23] F. Chung, R. Graham, and R. Wilson, Maximum cuts and quasi-random graphs, Random Graphs, (Poznan Conf., 1989) Wiley-Intersci, Publ. vol 2, 151–196.
- [24] D. Conlon and T. Gowers, Combinatorial theorems in sparse random sets, arXiv:1011.4310v1 [math.CO].
- [25] D. Daykin and R. Häggkvist, Degrees giving independent edges in a hypergraph, Bull. Austral. Math. Soc. 23 (1981), 103–109.
- [26] D. Dellamonica, Y. Kohayakawa, M. Marciniszyn, and A. Steger, On the resilience of long cycles in random graphs, *Electron. J. Combin.* 15 (2008), R32.
- [27] G. Dirac, Some theorems of abstract graphs, Proc. London Math. Soc. 3 (1952), 69–81.
- [28] P. Erdős, A problem on independent r-tuples, Ann. Univ. Sci. Budapest. Eötvös Sect. Math. 8 (1965), 93–95.
- [29] P. Erdős and T. Gallai, On the maximal paths and circuits of graphs, Acta Math. Acad. Sci. Hung., 10 (1959), 337–357.
- [30] P. Erdős, C. Ko and R. Rado, Intersection theorems for systems of finite sets, Quart. J. Math. Oxford Ser. (2), 12 (1961), 313–318.
- [31] S. Fajtlowicz. On conjectures of Graffiti, II, Congresus Numeratum 60 (1987), 189– 198.
- [32] U. Feige, On sums of independent random variables with unbounded variance and estimating the average degree in a graph, SIAM J. Comput. 35 (2006), no. 4, 964– 984.
- [33] J. Fox, P. Keevash and B. Sudakov, Directed graphs without short cycles, Combin. Probab. Comput., 19 (2010), 285–301.
- [34] P. Frankl, The shifting techniques in extremal set theory, in: Surveys in Combinatorics, Lond. Math. Soc. Lect. Note Ser. 123 (1987), 81–110.

- [35] P. Frankl and V. Rödl, Near perfect coverings in graphs and hypergraphs, *European J. Combin.*, 6(4) (1985), 317–326.
- [36] P. Frankl, V. Rödl, and A. Ruciński, On the maximum number of edges in a triple system not containing a disjoint family of a given size, submitted.
- [37] A. Frieze and M. Krivelevich, On two Hamilton cycle problems in random graphs, *Israel J. Math.* 166 (2008), 221–234.
- [38] Z. Gao, B. McKay, R. Naserasr, and B. Stevens, On Alon-Saks-Seymour Conjecture, *To appear*.
- [39] S. Gerke and A. Steger, The sparse regularity lemma and its applications, In Surveys in combinatorics 2005, volume 327 of London Math. Soc. Lecture Note Ser., Cambridge Univ. Press, Cambridge (2005), 227–258.
- [40] D. Gottlieb, A class of incidence matrices, Proc. Amer. Math. Soc. 17 (1966), 1233– 1237.
- [41] R. Graham and H. Pollak, On embedding graphs in squashed cubes. In Graph theory and applications, 303, Lecture Notes in Math., Springer, Berlin, 1972, 99–110.
- [42] A. Hajnal and E. Szemerédi, Proof of a conjecture of P. Erdős, In Combinatorial theory and its applications, II (Proc. Colloq., Balatonfüred, 1969), North-Holland, Amsterdam (1970), 601–623.
- [43] H. Hán, Y. Person and M. Schacht, On perfect matchings in uniform hypergraphs with large minimum vertex degree, SIAM J. Discrete Math. 23 (2009), no. 2, 732–748.
- [44] S. He, J. Zhang and S. Zhang, Bounding probability of small deviation: a fourth moment approach, Math. Oper. Res. 35 (2010), no.1, 208–232.
- [45] A. Hilton and E. Milner, Some intersection theorems for systems of finite sets, Quart. J. Math. Oxford Ser. (2), 18 (1967), 369–384.
- [46] W. Hoeffding, Probability inequalities for sums of bounded random variables, J. Ameri. Statist. Assoc., 58 (1963), 13–30.
- [47] H. Huang, P. Loh, and B. Sudakov, The size of a hypergraph and its matching number, *Combin. Probab. Comput.*, to appear.
- [48] S. Janson, Quasi-random graphs and graph limits, arXiv:0905.3241 [math.CO].
- [49] S. Janson, T. Łuczak, and A. Ruciński, *Random Graphs*, John Wiley and Sons, New York (2000).
- [50] J. Kahn, Coloring nearly-disjoint hypergraphs with n+o(n) colors, J. Combin. Theory Ser. A 59(1) (1992), 31–39.
- [51] J. Kahn, Recent results on some not-so-recent hypergraph matching and covering problems, in: *Extremal problems for finite sets (Visegrád, 1991)*, volume 3 of *Bolyai Soc. Math. Stud.*, János Bolyai Math. Soc., Budapest, 1994, 305–353.

- [52] I. Khan, Perfect matching in 3-uniform hypergraphs with large vertex degree, submitted.
- [53] I. Khan, Perfect matchings in 4-uniform hypergraphs, submitted.
- [54] D. Kleitman, On a conjecture of Milner on k-graphs with non-disjoint edges, J. Combinatorial Theory, 5 (1968), 153–156.
 J. Komlós, Tiling Turán theorems, Combinatorica 20 (2000), 203–218.
- [55] J. Komlós and M. Simonovits, Szemerédi's regularity lemma and its applications in graph theory, In Combinatorics, Paul Erdős is eighty, Vol. 2 (Keszthely, 1993), volume 2 of Bolyai Soc. Math. Stud., pages 295–352. János Bolyai Math. Soc., Budapest, 1996.
- [56] M. Krivelevich and B. Sudakov, Pseudo-random graphs, More sets, graphs and numbers, E. Györi, G. O. H. Katona and L. Lovász, Eds., Bolyai Society Mathematical Studies Vol. 15, 199–262.
- [57] D. Kühn and D. Osthus, Embedding large subgraphs into dense graphs Surveys in Combinatorics (editors S. Huczynka, J. Mitchell, C. Roney-Dougal), London Math. Soc. Lecture Notes, Cambridge University Press 365 (2009), 137–167.
- [58] D. Kühn, D. Osthus, and A. Treglown, Matchings in 3-uniform hypergraphs, submitted.
- [59] E. Kushilevitz, N. Linial, and R. Ostrovsky, The linear-array conjecture in communication complexity is false, *Combinatorica* **19**(2) (1999), 241–254.
- [60] E. Kushilevitz and N. Nisan, **Communication complexity**, Cambridge University Press, Cambridge, 1997.
- [61] C. Leiserson and J. Saxe, Retiming synchronous circuitry, Algorithmica 6 (1991), 5–35.
- [62] D. Leong, A. G. Dimakis, and T. Ho, Symmetric allocations for distributed storage, Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM), (2010).
- [63] D. Leong, A. G. Dimakis, and T. Ho, Distributed storage allocations, available at http://arxiv.org/pdf/1011.5287v1.pdf.
- [64] A. Lo and K. Markström, *F*-factors in hypergraphs via absorption, submitted.
- [65] L. Lovász, *Combinatorial problems and exercises*, North-Holland Publishing Co., Amsterdam, 1979.
- [66] L. Lovász and M. Saks, Lattices, Mobius functions, and communication complexity, Journal of Computer and System Sciences 47 (1993), 322–349.
- [67] N. Manickam, On the distribution invariants of association schemes, Ph.D. Dissertation, Ohio State University, 1986.

- [68] N. Manickam and D. Miklós, On the number of non-negative partial sums of a non-negative sum, Colloq. Math. Soc. Janos Bolyai 52 (1987), 385–392.
- [69] N. Manickam and N. M. Singhi, First distribution invariants and EKR theorems, J. Combinatorial Theory, Series A 48 (1988), 91–103.
- [70] G. Marino and G. Chiaselotti, A method to count the positive 3-subsets in a set of real numbers with non-negative sum, *European J. Combin.* **23** (2002), 619–629.
- [71] K. Markström and A. Ruciński, Perfect matchings and Hamilton cycles in hypergraphs with large degrees, *Europ. J. Combin.*, **32**(5) (2011), 677–687.
- [72] C. McDiarmid, Concentration, Probabilistic Methods for Algorithmic Discrete Mathematics (1998), 1–46.
- [73] D. Mubayi and S. Vishwanathan, Biclique Coverings and the Chromatic Number. *The Electronic Journal of Combinatorics* **16**(1) (2009).
- [74] M. Naor and R. M. Roth, Optimal file sharing in distributed networks, SIAM J. Computing, 24 (1991), 158–183.
- [75] M. Nathanson, The Caccetta-Häggkvist conjecture and additive number theory, AIM Preprint 2006-10: www.aimath.org/preprints.html.
- [76] N. Nisan and A. Wigderson, On rank vs. communication complexity, *Combinatorica* 15(4) (1995), 557–565.
- [77] G. W. Peck, A new proof of a theorem of Graham and Pollak, Discrete Math. 49(3) (1984), 327–328.
- [78] O. Pikhurko, Perfect matchings and K_4^3 -tilings in hypergraphs of large codegree, *Graphs Combin.* **24**(4) (2008), 391–404.
- [79] L. Pyber, A new generalization of the Erdős-Ko-Rado theorem, J. Combin. Theory Ser. A, 43 (1986), no. 1, 85–90.
- [80] R. Raz and B. Spieker, On the "log rank"-conjecture in communication complexity, *Combinatorica* **15**(4) (1995), 567–588.
- [81] A. A. Razborov, The gap between the chromatic number of a graph and the rank of its adjacency matrix is superlinear, *Discrete Math.* 108(1-3) (1992), 393–396.
- [82] V. Rödl and A. Ruciński, Dirac-type questions for hypergraphs a survey (or: more problems for Endre to solve), Bolyai Soc. Math. Studies 21 (2010), An Irregular Mind (Szemerédi is 70), Bolyai Soc. Math. Studies 21 (2010), 561–590.
- [83] V. Rödl, A. Ruciński, and E. Szemerédi, Perfect matchings in uniform hypergraphs with large minimum degree, *Europ. J. Combin.* 27 (2006) 1333–1349.
- [84] V. Rödl, A. Ruciński, and E. Szemerédi, An approximate Dirac-type theorem for kuniform hypergraphs, Combinatorica 28(2) (2008), 229–260.

- [85] V. Rödl, A. Ruciński, and E. Szemerédi, Perfect matchings in large uniform hypergraphs with large minimum collective degree, J. Combin. Theory, Ser. A 116 (2009), 613–636.
- [86] S. Samuels, On a Chebyshev-type inequality for sums of independent random variables, Ann. Math. Statist. **37** (1966), 248–259.
- [87] S. Samuels, More on a Chebyshev-type inequality for sums of independent random variables, Purdue Stat. Dept. Mimeo. Series no. 155 (1968)
- [88] M. Sardari, R. Restrepo, F. Fekri, and E. Soljanin, Memory allocation in distributed storage networks, *Proceedings of IEEE International Symposium on Information Theory (ISIT)*, (2010).
- [89] M. Schacht, Extremal results for random discrete structures, manuscript.
- [90] A. Shapira and R. Yuster, The quasi-randomness of hypergraph cut properties, arXiv:1002.0149v1 [math.CO].
- [91] A. Shaw, The logical design of operating systems, Prentice-Hall, Englewood Cliffs, NJ, 1974.
- [92] A. Shokoufandeh and Y. Zhao, Proof of a conjecture of Komlós, Random Structures Algorithms, 23 (2003), 180–205.
- [93] A. Shokoufandeh and Y. Zhao, On a tiling conjecture for 3-chromatic graphs, Discrete Math. 277 (2004), 171–191.
- [94] M. Simonovits and V. T. Sós, Hereditarily extended properties, quasi-random graphs and not necessarily induced subgraphs, *Combinatorica*, 17 (1997), 577–596.
- [95] B. Sullivan, Extremal problems in digraphs, PhD thesis, Princeton University, 2008.
- [96] B. Sullivan, A summary of results and problems related to the Caccetta-Häggkvist conjecture, available online at http://arxiv.org/abs/math/0605646v1.
- [97] E. Szemerédi, Regular partitions of graphs, In Problèmes combinatoires et théorie des graphes (Colloq. Internat. CNRS, Univ. Orsay, Orsay, 1976), volume 260 of Colloq. Internat. CNRS, pages 399–401. CNRS, Paris, 1978.
- [98] A. Thomason, *Pseudo-random graphs*, in: Proceedings of Random Graphs, Poznań 1985, M. Karoński, ed., Annals of Discrete Math. **33** (North Holland 1987), 307–331.
- [99] A. Thomason, Random graphs, strongly regular graphs and pseudo-random graphs, Surveys in Combinatorics, 1987, C. Whitehead, ed., LMS Lecture Note Series 123 (1987), 173–195.
- [100] H. Tverberg, On the decomposition of K_n into complete bipartite graphs, J. Graph Theory **6**(4) (1982), 493-494.

- [101] M. Tyomkyn, An improved bound for the Manickam-Miklós-Singhi conjecture, available online at http://arxiv.org/pdf/1011.2803v1.
- [102] C. Van Nuffelen, Rank, Clique, and Chromatic Number of a Graph, System Modeling and Optimization, 38, Lect. Notes Control Inf. Sci, 605–611.
- [103] S. Vishwanathan, A polynomial space proof of the Graham-Pollak theorem, J. Combin. Theory Ser. A 115(4) (2008), 674–676.
- [104] M. Yannakakis, Expressing combinatorial optimization problems by linear programs, Journal of Computer and System Sciences 43(3) (1991), 441–466.
- [105] A. Yao, Some complexity questions related to distributive computing, Proceedings of the eleventh annual ACM Symposium on Theory of Computing, ACM New York (1979), 209–213.
- [106] J. Zaks. Bounds of neighborly families of convex polytopes. Geometriae Dedicata 8(3) (1979), 279–296.