Daniel M. Martin

Locally nearly perfect packings

By

Daniel M. Martin
Doctor of Philosophy

Mathematics

Vojtěch Rödl, Ph.D.
Advisor

Ron Gould, Ph.D.
Committee Member

Robert Roth, Ph.D.
Committee Member

Accepted:

Lisa A. Tedesco, Ph.D.
Dean of the Graduate School

Date

Locally nearly perfect packings

by

Daniel M. Martin
M.S. in Computer Science, University of São Paulo, 2005

Advisor: Vojtěch Rödl, Ph.D.

Abstract of
A dissertation submitted to the Faculty of the Graduate School
of Emory University in partial fulfillment
of the requirements of the degree of
Doctor of Philosophy

Mathematics

2009

Abstract

Locally nearly perfect packings
By Daniel M. Martin

In 1963 P. Erdős and H. Hanani conjectured that for every fixed positive integers $\ell < k$, and for every n there exists a family $\mathcal{F}_n$ of k-element subsets of $\{1, 2, \ldots, n\}$ with the property that every ℓ-element subset of $\{1, 2, \ldots, n\}$ is contained in at most one member of $\mathcal{F}_n$, and the proportion of ℓ-element subsets which are not contained in any member of $\mathcal{F}_n$ tends to 0 as n tends to infinity. The conjecture was solved by V. Rödl in 1985. Since then, many improvements and generalizations followed Rödl's result. Until the present moment all known proofs of the conjecture use the "semi-random method", an iterative process in which the desired set system is build as a successive union of small randomly chosen pieces. In the first part of this dissertation we give an alternative proof of Rödl's result using a somewhat different approach which is elementary and fairly simple. In the second part, we use the semi-random method to show a strengthening of Rödl's result in which the family $\mathcal{F}_n$ is also required to satisfy that, for every $0 < j < \ell$ and for all j-element subset $J \subseteq \{1, 2, \ldots, n\}$, the proportion of ℓ-element subsets containing J which are not contained in any member of $\mathcal{F}_n$ tends to 0 as n tends to infinity. This means that the family $\mathcal{F}_n$ is close to being a perfect packing not only globally, but also locally.

Locally nearly perfect packings

by

Daniel M. Martin
M.S. in Computer Science, University of São Paulo, 2005

Advisors: Vojtěch Rödl, Ph.D.

A dissertation submitted to the Faculty of the Graduate School
of Emory University in partial fulfillment
of the requirements of the degree of
Doctor of Philosophy

Mathematics

2009

Acknowledgments

I thank the Lord my God.

I thank my supervisor Professor Vojtěch Rödl for his patience and incredible help. This dissertation would not have come to existence without his constant advices. I am also grateful to the committee members.

I appreciate the great support that came from my entire family, and I thank Professor Yoshiharu Kohayakawa which greatly contributed to my growth as a mathematician.

I am grateful for Jan Zich who introduced me to many of the techniques applied in the second half of this dissertation. Special thanks to Christian, Colton, and Eduardo.

Finally, I express my gratitude to all my friends who have been a constant blessing throughout the past years of my live. May our friendship last forever.

To Thaís

Contents

List of Tables

Chapter 1

Introduction

1.1 A problem on combinations

In the fields of combinatorics and design theory, there is an interesting class of objects which has been studied extensively over the last 150 years. These objects are named *Steiner systems*. We introduce the reader to Steiner systems with a question that appeared in the 1844 issue of an almanac published in London called *Lady's and Gentleman's Diary*[1] devoted to mathematical problems and enigmas.

> *"Determine the number of combinations that can be made out of* n *symbols, each combination having* p *symbols, with this limitation, that no combination of* q *symbols which may appear in any one of them, may be repeated in any other."*

Since the general problem turned out to be too difficult for the readers, a later issue specialized the question to the case $q = 2, p = 3$. A few years later, the British mathematician T. P. Kirkman solved this case, and published his solution in the second volume of the *Cambridge and Dublin Mathematical Journal* [14].

[1] The *Lady's and Gentleman's Diary* is a successor of the periodical *Ladies Diary*. See [1] for a compendium of its contributions to mathematical and exact sciences.

We restate the problem in a different language. For conciseness, denote the set of the first n positive integers $\{1, 2, \ldots, n\}$ by $[n]$.

Problem 1.1 *Given positive integers $\ell < k < n$, what is the largest cardinality of a family of k-element subsets of $[n]$ with the property that any ℓ-element subset is contained in at most one member of the family?*

Below, an easy upper bound for the size of such families is given.

Proposition 1.2 *Let $\mathcal{F}$ be a family of k-subsets of $[n]$ with the property that no ℓ-subset is contained in more than one member of $\mathcal{F}$. Then*

$$|\mathcal{F}| \leq \binom{n}{\ell}\binom{k}{\ell}^{-1}. \tag{1.1}$$

Proof. We count pairs (L, K) with $L \subseteq K$, where L is an ℓ-subset of $[n]$, and K belongs to $\mathcal{F}$. Note that each member K of $\mathcal{F}$ appears in exactly $\binom{k}{\ell}$ pairs. Therefore the total number of pairs is $|\mathcal{F}|\binom{k}{\ell}$. Since each ℓ-subset is contained in at most one member of $\mathcal{F}$, it is also counted in at most one pair. Hence, the total number of pairs must not exceed the total number of ℓ-subsets, which implies

$$|\mathcal{F}|\binom{k}{\ell} \leq \binom{n}{\ell}.$$

Moreover, equality holds if and only if each ℓ-subset belongs to precisely one member of $\mathcal{F}$. The proposition follows. □

A family $\mathcal{F}$ satisfying the hypothesis of Proposition 1.2 is called a *partial Steiner (n, k, ℓ)-system*, or an *(n, k, ℓ)-system* for short. Alternatively, $\mathcal{F}$ may be called *packing*, because the ℓ-subsets are being "packed" into k-subsets. Furthermore, if $\mathcal{F}$ satisfies (1.1) with equality, then $\mathcal{F}$ is a *Steiner (n, k, ℓ)-system.* The special case of Steiner $(n, 3, 2)$-systems are traditionally called *Steiner triple systems.*

1.1.1 Steiner systems

Even though the study of such set systems were initiated by Kirkman, they were named after the Swiss mathematician J. Steiner who posed questions related to Problem 1.1 in the *Crelle's Journal fur Mathematik*, vol. 45 (1853). In 1859, M. Reiss [19] solved one of Steiner's questions which corresponded precisely to the case $\ell = 2, k = 3$ of Problem 1.1. His achievement was independent of Kirkman, and in fact, it took many years until the mathematical community gave Kirkman credit for his contributions on this problem.

Existence of Steiner Systems

In this section we intend to briefly survey what is known on the existence of Steiner systems. We start with a necessary and sufficient condition for the occurrence of Steiner triple systems.

Proposition 1.3 *There exists a Steiner triple system on $n \geq 1$ elements if and only if $n \equiv 1, 3 \pmod 6$.*

For a proof, see Chapter 8 in [6]. The necessity part of Proposition 1.3 is actually a consequence of the following general criteria given by H. Hanani in [10].

Proposition 1.4 *If there exists a Steiner (n, k, ℓ)-system on $n \geq 1$ elements, then $\binom{k-j}{\ell-j}$ divides $\binom{n-j}{\ell-j}$ for every $j = 0, 1, \ldots, \ell - 1$.*

Proof. Let $\mathcal{F}$ be an (n, k, ℓ)-system on n elements. Let $0 \leq j < \ell$ be fixed, and let J be a subset of $[n]$ of cardinality j. We restrict our attention to the ℓ-subsets of $[n]$ containing J. The number of ℓ-subsets of interest is $\binom{n-j}{\ell-j}$, each of which is contained in precisely one member of $\mathcal{F}$. In turn, each $K \in \mathcal{F}$ with $J \subseteq K$, contains $\binom{k-j}{\ell-j}$ ℓ-subsets of interest. Therefore, it must be the case that $\binom{k-j}{\ell-j}$ divides $\binom{n-j}{\ell-j}$. Since j was arbitrary, the proposition follows. □

In general, given k and ℓ, the problem of deciding whether Steiner (n, k, ℓ)-systems exist is still wide open. The criteria given by Proposition 1.4 is shown [10, 11] to be sufficient for $(k, \ell) = (3, 2), (4, 2), (5, 2), (4, 3)$. Thanks to a sequence of difficult papers by R. M. Wilson [23, 24, 25], it is also known to be sufficient for $\ell = 2$, any k, and $n \geq n_0(k)$. On the other hand, the criteria is not always sufficient. For example $(n^2 + n + 1, n + 1, 2)$-Steiner Systems are equivalent to projective planes of order n. The divisibility conditions in Proposition 1.4 are satisfied, but a famous theorem of R. H. Bruck and H. J. Ryser [5] gives an infinite class of numbers n for which projective planes do not exist.

There are some infinite families of Steiner systems which are known to exist. The *Handbook of Combinatorial Designs* [7] lists examples of Steiner systems and a few inductive constructions. We summarize all in Table 1.1 in the next page. For conciseness, instead of writing "a Steiner (n, k, ℓ)-system exists", we write "(n, k, ℓ) exists".

Table 1.1: Steiner systems known to exist

n	k	ℓ	Conditions
q^n	q	2	q is a prime power, and $n \geq 2$
$q^n + \cdots + q + 1$	$q+1$	2	q is a prime power, and $n \geq 2$
$q^3 + 1$	$q+1$	2	q is a prime power
$2^{r+s} + 2^r - 2^s$	2^r	2	$2 \leq r < s$
$q^n + 1$	$q+1$	3	q is a prime power, and $n \geq 2$
$4t - 2$	6	3	provided $(t, 6, 3)$ exists
$rs + 1$	$q+1$	3	q is a prime power, and both $(r+1, q+1, 3)$ and $(s+1, q+1, 3)$ exist
n	k	ℓ	provided $(n+1, k+1, \ell+1)$ exists

On the other hand, only finitely many examples of Steiner systems are known for $\ell \geq 4$. An even more surprising fact is that no Steiner system is known to exist for $\ell \geq 6$.

1.1.2 Partial Steiner systems

Let us denote by $S(n, k, \ell)$ the maximum cardinality of a partial Steiner (n, k, ℓ)-system. In view of the discussion in the preceding section, finding an answer to the general question in the *Lady's and Gentleman's Diary*, i.e. computing the value of $S(n, k, \ell)$ for any $\ell < k < n$, seems completely hopeless. For this reason, we turn our attention to a related (although still very difficult) problem which has already been solved.

1.2 The conjecture of P. Erdős and H. Hanani

The following is a very natural question, which perhaps has already crossed the mind of the reader. How far is $S(n, k, \ell)$ from the upper bound given in Proposition 1.2? Can we find a good lower bound? In fact, P. Erdős and H. Hanani [8] conjectured in 1963 that for every fixed k and ℓ, and n tending to infinity, there should be (n, k, ℓ)-systems which are really close to being Steiner. In other words, they conjectured that, as n goes to infinity, it is possible to have packings with only a $o(1)$-proportion of all the ℓ-subsets of $[n]$ being uncovered. Such packings are called nearly perfect. The precise statement of the conjecture is written below.

Conjecture 1.5 (Erdős and Hanani, 1963) *For fixed integers* $0 < \ell < k$ *the following holds*

$$\lim_{n\to\infty} \frac{S(n,k,\ell)}{\binom{n}{\ell}/\binom{k}{\ell}} = 1. \tag{1.2}$$

In [16], N. Kuzjurin studied a variant of this problem in which k is not constant, but varies with n. He considered an algebraic construction, which was first introduced by Zinoviev [26] in 1965, to show that, if $k = k(n)$ tends to infinity, $\ell = \ell(n) = o(k(n))$, and $k(n) < c\sqrt{n}$ for some $c < 1$, then (1.2) holds. On the other hand, he proved that nearly perfect packings do not exist if $k > c\sqrt{n}$ for some constant $c > 1$.

In 1985, V. Rödl [20] proved that Conjecture 1.5 is indeed true. In his seminal paper, he introduced a technique which is called today the *semi-random method* or the *Rödl nibble*. This paper caught the attention of several authors who were interested both in the beauty of the problem and in the ingenuity of the proof. Many of them further improved and generalized Rödl's result in different ways. These authors all seem to benefit from an important link that was first observed by P. Frankl and V. Rödl in [9] between the problem of finding nearly perfect packings and the problem of finding large matchings in hypergraphs. In the next section, we give some definitions and explain the connection between these two problems.

1.2.1 A connection with matchings in hypergraphs

A *hypergraph* is an ordered pair $(V, \mathcal{H})$, where V is a finite set and $\mathcal{H}$ is a family of subsets of V. Elements of V are called *vertices*, members of $\mathcal{H}$ are called *edges*. Sometimes we refer to $(V, \mathcal{H})$ simply as $\mathcal{H}$. A hypergraph is said to be *uniform* if all edges have the same cardinality. A *matching* in a hypergraph $\mathcal{H}$ is simply a subfamily $\mathcal{M} \subseteq \mathcal{H}$ with the property that no two members of $\mathcal{M}$ intersect. Define the *degree* $\deg(v)$ of a vertex v to be the number of members of $\mathcal{H}$ containing v, and the *codegree* $\deg(u, v)$ to be the number of members containing both u and v. If all degrees are the same, $\mathcal{H}$ is said to be *regular*. The minimum and maximum degree of $\mathcal{H}$ are denoted by $\delta(\mathcal{H})$ and $\Delta(\mathcal{H})$ respectively. The maximum codegree of $\mathcal{H}$ is denoted by $\Delta_2(\mathcal{H})$. For a set X and a integer k, let $\binom{X}{k}$ denote the family of all k-element subsets of X.

As mentioned earlier, P. Frankl and V. Rödl were the first to point out a connection between the problem of finding large (n, k, ℓ)-systems and the problem of finding large matchings in hypergraphs.

The connection is described as follows. Suppose $\ell < k < n$ are positive integers. Set $V = \binom{[n]}{\ell}$, i.e. elements of V are ℓ-subsets of $[n]$. Consider the hypergraph $(V, \mathcal{H})$ defined by

$$\mathcal{H} = \mathcal{H}(n, k, \ell) = \{\tbinom{K}{\ell} : K \in \tbinom{[n]}{k}\}.$$

A subset of V constitutes an edge of $\mathcal{H}$ if and only if the corresponding ℓ-subsets are all the ℓ-subsets contained in some k-subset of $[n]$. Hence, edges of $\mathcal{H}$ are in one-to-one correspondence with k-subsets of $[n]$. Note that $\mathcal{H}$ has $\binom{n}{\ell}$ vertices, it is $\binom{k}{\ell}$-uniform, and $\binom{n-\ell}{k-\ell}$-regular.

Now consider an (n, k, ℓ)-system $\mathcal{F}$. Since two members of $\mathcal{F}$ do not share ℓ elements, the edges of $\mathcal{H}$ corresponding to members of $\mathcal{F}$ form a matching in $\mathcal{H}$. Conversely, if $\mathcal{M}$ is a matching in $\mathcal{H}$, then the k-subsets of $[n]$ corresponding to the edges in $\mathcal{M}$ form an (n, k, ℓ)-system. That establishes the correspondence between (n, k, ℓ)-systems and matchings of the hypergraph $\mathcal{H}(n, k, \ell)$.

Now let $\mathcal{M}$ be a matching of $\mathcal{H}(n, k, \ell)$, and let $\mathcal{F}$ be the corresponding (n, k, ℓ)-system. Suppose $\mathcal{M}$ covers all but an ε-fraction of the vertices in V. Then the proportion of ℓ-subsets which are not covered by any member of $\mathcal{F}$ is also ε, which implies

$$|\mathcal{F}| \geq (1-\varepsilon)\binom{n}{\ell}\binom{k}{\ell}^{-1}. \tag{1.3}$$

1.2.2 Exploiting the connection

The paper of Frankl and Rödl [9] also contained a generalization to Rödl's first proof of Conjecture 1.5. We omit their result, which was later strengthened by N. Pippenger (unpublished) as follows.

Theorem 1.6 *For an integer $r \geq 2$, and a real $\varepsilon > 0$ there exists a real $\mu = \mu(r, \varepsilon)$ such that the following holds. If the r-uniform hypergraph $\mathcal{H}$ on n vertices satisfies*

(i) $\delta(\mathcal{H}) \geq (1-\mu)\Delta(\mathcal{H})$,

(ii) $\Delta_2(\mathcal{H}) < \mu\Delta(\mathcal{H})$,

then $\mathcal{H}$ has a perfect matching that covers all but at most εn vertices.

Proposition 1.7 *Theorem 1.6 implies Conjecture 1.5.*

Proof. Fix positive integers $\ell < k$. Set $r = \binom{k}{\ell}$. Let $\varepsilon > 0$ be arbitrary, and get $\mu = \mu(r, \varepsilon)$ from Theorem 1.6. We want to show that, for sufficiently large n, the r-uniform hypergraph $\mathcal{H} = \mathcal{H}(n, k, \ell)$ satisfies both conditions of Theorem 1.6. Condition (i) is clearly satisfied, since $\mathcal{H}$ is regular of degree $\binom{n-\ell}{k-\ell} = O(n^{k-\ell})$. To verify that $\mathcal{H}$ satisfies condition (ii), note that $\Delta_2(\mathcal{H})$ corresponds to the number of k-subsets of $[n]$ that contain two fixed distinct ℓ-subsets of $[n]$. Since any two ℓ-subsets span at least $\ell+1$ elements, $\Delta_2(\mathcal{H}) = \binom{n-\ell-1}{k-\ell-1} = o(n^{k-\ell})$. Therefore, for sufficiently large n, condition (ii) holds. Consequently, by Theorem 1.6, $\mathcal{H}$ has a matching $\mathcal{M}$ that covers all but εn vertices. As previously discussed in Section 1.2.1, the (n, k, ℓ)-system corresponding to $\mathcal{M}$ must satisfy inequality (1.3). Since ε is arbitrary, Conjecture 1.5 follows. $\square$

Various authors have considered the problem of finding large matchings in nearly regular hypergraphs. We cite, for example, N. Pippenger and J. Spencer [18], N. Alon, J. Kim and J. Spencer [2], A. Kostochka, V. Rödl and L. Talysheva [15], V. Vu [22], N. Alon and R. Yuster [4].

Their results can also be stated in terms of finding nearly perfect (n, k, ℓ)-systems. Some of these results improve the fraction of covered ℓ-subsets, while others prove the existence of set systems satisfying stronger properties. The original result of Rödl, however, gives no upper bound on the fraction of uncovered ℓ-subsets. This is also the case of [18] and [4] although these papers are strong generalizations of [20]. The author who currently has the best known upper bound on the fraction of uncovered ℓ-subsets is V. Vu [22]. For the general case, he proved the existence of (n, k, ℓ)-systems with only $O(n^{\ell-\beta} \ln^{\gamma} n)$ uncovered ℓ-subsets, where $\beta = 1/\left(\binom{k}{\ell} - 1\right)$, and $\gamma > 0$. For $k > \ell+3$, he gave even a better bound, namely $O(n^{\ell-\beta(k-\ell)/3} \ln^{\gamma} n)$, for some other $\gamma > 0$.

1.3 A result of interest

We have cited both [2] and [4] as generalizations of [20]. Since this dissertation is especially motivated by these two papers, we find it relevant to mention the first briefly, and to discuss the second in more details.

The paper of N. Alon, J. Kim and J. Spencer [2] is very important to us, because we adapt many of its techniques, and use them in the proof of the main lemma of Chapter 4. Another nice feature of [2] is the introduction of a martingale inequality which we apply a couple of times (see Section 2.2.3 in Chapter 2 for the statement of this inequality).

We now describe the main result of N. Alon and R. Yuster in [4] and point out one of its consequences. Let $(V, \mathcal{H})$ be a uniform hypergraph, and let $\mathcal{F} \subseteq 2^V$ be a family of subsets of V. A matching $\mathcal{M}$ in $\mathcal{H}$ is called $(\alpha, \mathcal{F})$-perfect if, for each $F \in \mathcal{F}$, at least $\alpha|F|$ vertices of F are covered by $\mathcal{M}$. The main result in [4] gives sufficient conditions for the existence of a $(1-\varepsilon, \mathcal{F})$-perfect matching in a uniform hypergraph.

Before we state their result, let us introduce some more notation. For $\mathcal{F} \subseteq 2^V$, let $s(\mathcal{F})$ denote $\min_{F \in \mathcal{F}} |F|$. Given a hypergraph $\mathcal{H}$, let $g(\mathcal{H}) = \Delta(\mathcal{H})/\Delta_2(\mathcal{H})$.

Theorem 1.8 (N. Alon and R. Yuster, 2005) *For an integer $r \geq 2$, a real $C > 1$ and a real $\varepsilon > 0$ there exists a real $\mu = \mu(r, C, \varepsilon)$ and a real $K = K(r, C, \varepsilon)$ such that the following holds. If an r-uniform hypergraph on n vertices satisfies*

(i) $\delta(\mathcal{H}) \geq (1-\mu)\Delta(\mathcal{H})$, *and*

(ii) $g(H) > \max\{1/\mu, K \ln^6 n\}$,

then for every $\mathcal{F} \subseteq 2^V$ with $|\mathcal{F}| \leq C^{g(\mathcal{H})^{1/(3r-3)}}$, and with $s(\mathcal{F})$ being at least $5g(\mathcal{H})^{1/(3r-3)} \ln(|\mathcal{F}| g(\mathcal{H}))$, there is a $((1-\varepsilon), \mathcal{F})$-perfect matching in $\mathcal{H}$.

Essentially, the hypothesis of Theorem 1.8 require the hypergraph to be nearly regular, the maximum codegree to be relatively small compared to the maximum degree, and the family $\mathcal{F}$ to satisfy two modest constraints: an upper bound on its cardinality, and a lower bound on the cardinality of its smallest member.

Alon and Yuster use Theorem 1.8 to generalize Rödl's result in the following sense.

Corollary 1.9 *Let $1 < \ell < k$ be fixed integers, and let $\varepsilon > 0$. For n sufficiently large there exists an (n, k, ℓ)-system such that every element of $[n]$ belongs to at most $\varepsilon\binom{n-1}{\ell-1}$ uncovered ℓ-subsets.*

What Corollary 1.9 says is that, for large enough n, it is possible not only to find packings which are nearly perfect, but also locally nearly perfect. By looking at the proof of Corollary 1.9 (i.e. Theorem 3.1 in [4]) we were able to further extend Corollary 1.9 as follows.

Corollary 1.10 *Let $1 < \ell < k$ be fixed integers, and let $\varepsilon > 0$. For n sufficiently large there exists an (n, k, ℓ)-system such that, for every j with $0 \leq j < \ell$, and for every j-subset $J \subseteq [n]$, the set J is contained in at most $\varepsilon\binom{n-j}{\ell-j}$ uncovered ℓ-subsets.*

Proof. We apply Theorem 1.8 with ε, $r = \binom{k}{\ell}$, and $C = 2$. Consider the hypergraph $\mathcal{H} = \mathcal{H}(n, k, \ell)$ from Section 1.2.1. Recall that $\mathcal{H}$ is $\binom{k}{\ell}$-uniform, and has $N = \binom{n}{\ell}$ vertices. Also, $\delta(\mathcal{H}) = \Delta(\mathcal{H}) = \binom{n-\ell}{k-\ell}$. This implies condition (i) in Theorem 1.8 is satisfied. Recall also that any two ℓ-subsets of $[n]$ span at least $\ell + 1$ points. Thus $\Delta_2(\mathcal{H}) = \binom{n-\ell-1}{k-\ell-1}$, and $g(\mathcal{H}) = (n-\ell)/(k-\ell) = \Theta(n)$, which guarantees the validity of condition (ii) in Theorem 1.8. For each subset $J \subseteq [n]$, with $0 \leq |J| < \ell$, let F_J be the family of ℓ-subsets of $[n]$ containing J. Note also that F_J is a subset of vertices of $\mathcal{H}$ with $|F_J| = \binom{n-j}{\ell-j}$, where $j = |J|$. Let $\mathcal{F} = \{F_J \colon J \subseteq [n], 0 \leq |J| < \ell\}$. Thus

$$|\mathcal{F}| = \sum_{j=0}^{\ell-1} \binom{n-j}{\ell-j} = \Theta(n^\ell),$$

and $s(\mathcal{F}) = (n - \ell + 1) = \Theta(n)$.

We need to verify that the remaining conditions of Theorem 1.8 are satisfied. For n sufficiently large (and consequently also N large) we have

$$C^{g(\mathcal{H})^{1/(3r-3)}} \geq C^{\Theta(n^{1/(3r-3)})} > \Theta(n^{\ell}) = |\mathcal{F}|.$$

Also, $s(\mathcal{F}) = \Theta(n) > 5\Theta(n^{1/(3r-3)})\Theta(\ln n) = 5g(\mathcal{H})^{1/(3r-3)}\ln(|\mathcal{F}|g(\mathcal{H}))$. Therefore, $\mathcal{H}$ has a $(1-\varepsilon)$-perfect matching. This, in turn, implies that there is an (n,k,ℓ)-system such that each j-subset, $0<j<\ell$ is contained in at most $\varepsilon\binom{n-j}{\ell-j}$ uncovered ℓ-subsets as desired. □

1.4 Main results

We are ready to state our two main results.

We start with the result in Chapter 4, because it is directly related to what we have been discussing in the previous section. The goal is to give a joint generalization of Corollary 1.10 and the result of V. Vu in [22]. We want to find locally nearly perfect packings, and further bound the proportion of uncovered ℓ-subsets as function that decreases polynomially in n. We were able to obtain the same bound as V. Vu gets in [22] for the general case. We did not achieve, however, the bound he obtains for the special case $k>\ell+3$. More specifically, we prove the following theorem.

Theorem 1.11 *Let $\ell<k$ be fixed. For every n there exists an (n,k,ℓ)-system such that, for every $0\leq j<\ell$, and for every j-subset $J\subseteq[n]$, the number of uncovered ℓ-subsets containing J is*

$$O(n^{\ell-j-\beta}\ln^{\gamma} n),$$

where $\beta=1/\left(\binom{k}{\ell}-1\right)$, and $\gamma>0$ is constant.

The proof of Theorem 1.11 also uses a variant of the semi-random method. In fact, no previous proof of Conjecture 1.5 (or of any result that implies it) is known that does not use this method. In contrast, our second main result is an elementary proof of Conjecture 1.5 that does not use the semi-random method. We use the same algebraic construction as Kuzjurin does in [16], together with a simple counting argument, to prove (1.2). The bound on the proportion of uncovered ℓ-subsets, however, is much weaker than the bound in [22]. We are only able to show that all but $O\big(\binom{n}{\ell}(\log\log n)^{-\beta}\big)$ ℓ-subsets are covered.

1.5 A word about packings and coverings

A problem which is the natural dual to the problem of finding maximum packings is the following. What is the minimum number $s(n,k,\ell)$ of k-element subsets of $[n]$ in a family $\mathcal{E}$ with the property that every ℓ-element subset of $[n]$ is contained in at least one member of $\mathcal{E}$? A family of k-sets as above is called an (n,k,ℓ)-*covering*, because every ℓ-subset of $[n]$ is "covered" by one of the members of the family. With a proof very much similar to that of Proposition 1.2, one has the following proposition, which provides a lower bound on $s(n,k,\ell)$.

Proposition 1.12 *Let $\ell < k < n$ be positive integers. Then*

$$s(n,k,\ell) \geq \binom{n}{\ell}\binom{k}{\ell}^{-1}.$$

Conjecture 1.5 was originally stated in [8] in a seemingly stronger form. Besides (1.2), it was also conjectured that the following identity is true.

$$\lim_{n\to\infty} \frac{s(n,k,\ell)}{\binom{n}{\ell}/\binom{k}{\ell}} = 1. \tag{1.4}$$

We decided to omit (1.4) from the statement of Conjecture 1.5 because it turns out that (1.4) and (1.2) are equivalent.

Proposition 1.13 (1.4) *and* (1.2) *are equivalent.*

Proof. The equivalence is very simple to demonstrate. First, suppose (1.2) is true. Fix $\varepsilon > 0$ and let $n_0 = n_0(\varepsilon)$ be such that, for all $n \geq n_0$, we have $S(n,k,\ell) \geq (1-\varepsilon/\binom{k}{\ell})\binom{n}{\ell}/\binom{k}{\ell}$. Let $\mathcal{F}$ be an optimal (n,k,ℓ)-system, i.e. $|\mathcal{F}| = S(n,k,\ell)$. An (n,k,ℓ)-covering $\mathcal{E}$ may be constructed from $\mathcal{F}$ as follows. For each uncovered ℓ-subset $L \subseteq [n]$, we add to the family some k-set containing L. The resulting family $\mathcal{E}$ is clearly a covering, and has at most $\varepsilon\binom{n}{\ell}/\binom{k}{\ell}$ additional members. Therefore

$$\begin{aligned} |\mathcal{E}| &\leq |\mathcal{F}| + \varepsilon\binom{n}{\ell}/\binom{k}{\ell} \\ &= (1-\varepsilon/\binom{k}{\ell})\binom{n}{\ell}\binom{k}{\ell}^{-1} + \varepsilon\binom{n}{\ell}/\binom{k}{\ell} \\ &= (1-\varepsilon/\binom{k}{\ell} + \varepsilon)\binom{n}{\ell}\binom{k}{\ell}^{-1} \\ &< (1+\varepsilon)\binom{n}{\ell}\binom{k}{\ell}^{-1}. \end{aligned}$$

Since $s(n,k,\ell) \leq |\mathcal{E}|$, and ε is arbitrary, (1.4) follows. The converse has a similar proof. $\square$

Chapter 2

Probabilistic tools

2.1 Basic probability

2.1.1 Probability spaces

A *finite probability space* consists of a finite set Ω together with a function $\Pr\colon \Omega \to [0,1]$ satisfying

$$\sum_{\omega \in \Omega} \Pr(\omega) = 1.$$

The set Ω is called the *sample space*, and Pr is called the *probability function*. A *trial* consists of sampling an element from Ω in accordance with the given probability function. The resulting element is called the *outcome* of the trial.

Subsets of the sample space Ω are called *events*. We can naturally extend Pr to all events as follows. For any $A \subseteq \Omega$, define $\Pr(A)$ as

$$\Pr(A) = \sum_{\omega \in A} \Pr(\omega).$$

In order to differentiate sets that are events from usual sets, we use $\wedge$ and $\vee$ to denote their intersection and union respectively. The complement of an event A is defined as $\overline{A} = \Omega \setminus A$.

In this dissertation, the definition of certain probability spaces implicitly depend on a parameter n, which goes to infinity. In this case, given an event A in such a probability space, we say that A happens *asymptotically almost surely* (*a.a.s.* for short) if $\Pr(A)$ tends to 1 as n tends to infinity.

2.1.2 Inclusion-Exclusion

We need the following tool in Chapter 4.

Proposition 2.1 (Inclusion-Exclusion Formula) *Let $A_1, \dots, A_k \subseteq \Omega$ be events in a probability space. For $I \subseteq [k]$, set $A_I = \bigwedge_{i \in I} A_i$, and $A_\emptyset = \Omega$. Then*

$$\sum_{I \subseteq [k]} (-1)^{|I|} \Pr(A_I) = 1 - \Pr\left(\bigvee_{i \in [k]} A_i\right).$$

In the case of finite probability spaces, a proof can be found in *Combinatorial Problems and Exercises* by L. Lovász [17] (see 2.2 part (b)).

2.1.3 Independence

Two events A and B are said to be *independent* if

$$\Pr(A \wedge B) = \Pr(A)\Pr(B).$$

Similarly, events $A_1, A_2, \dots, A_k$ are said to be *mutually independent* if for any $\mathcal{B} \subseteq \{A_1, A_2, \dots, A_k\}$ we have

$$\Pr\left(\bigwedge_{B \in \mathcal{B}} B\right) = \prod_{B \in \mathcal{B}} \Pr(B).$$

It is straightforward to verify that, given a set $\mathcal{A}$ of mutually independent events, and subsets $\mathcal{B}, \mathcal{C} \subseteq \mathcal{A}$ with $\mathcal{B} \cap \mathcal{C} = \emptyset$, the following holds.

$$\Pr\left(\bigwedge_{B \in \mathcal{B}} B \wedge \bigwedge_{C \in \mathcal{C}} \overline{C}\right) = \prod_{B \in \mathcal{B}} \Pr(B) \prod_{C \in \mathcal{C}} \Pr(\overline{C}).$$

2.1.4 Random variables

A *random variable* over Ω is simply a function $X\colon \Omega \to \mathbb{R}$.

Sometimes events can be described in terms of a random variable. For example, when we consider the event that $X > 17$, we are referring to the subset of Ω described by $\{\omega \in \Omega\colon X(\omega) > 17\}$.

The *expectation* of a random variable X, or the *expected value* of X, is defined to be

$$\mathbb{E}(X) = \sum_{\omega \in \Omega} X(\omega) \Pr(\omega).$$

An important property of the expectation of random variables is that it is a linear function. If X and Y are random variables over a certain probability space, then

$$\begin{aligned} \mathbb{E}(X+Y) &= \sum_{\omega \in \Omega} [X+Y](\omega) \Pr(\omega) \\ &= \sum_{\omega \in \Omega} \big(X(\omega) + Y(\omega)\big) \Pr(\omega) \\ &= \sum_{\omega \in \Omega} X(\omega) \Pr(\omega) + \sum_{\omega \in \Omega} Y(\omega) \Pr(\omega) \\ &= \mathbb{E}(X) + \mathbb{E}(Y). \end{aligned}$$

2.2 Concentration

2.2.1 Markov's Inequality

Let X be a non-negative random variable (i.e. $X(\omega) \geq 0$ for all $\omega \in \Omega$). Markov's Inequality can be stated as follows. For $t > 0$, the probability that $X \geq t$ is bounded by

$$\Pr(X \geq t) \leq \frac{\mathbb{E}(X)}{t}. \tag{2.1}$$

Clearly, this bound is only useful if $t > \mathbb{E}(X)$. Below, a proof is given.

$$\begin{aligned}\mathbb{E}(X) &= \sum_{\omega\in\Omega} X(\omega)\Pr(\omega)\\ &= \sum_{\substack{\omega\in\Omega\\ X(\omega)<t}} X(\omega)\Pr(\omega) + \sum_{\substack{\omega\in\Omega\\ X(\omega)\geq t}} X(\omega)\Pr(\omega)\\ &\geq \sum_{\substack{\omega\in\Omega\\ X(\omega)\geq t}} t\Pr(\omega)\\ &= t\Pr(X\geq t).\end{aligned}$$

Inequality (2.1) follows.

2.2.2 The Chernoff Bound

In Chapter 4, we need to work with random variables which can be expressed as a sum of others independent random variables, Chernoff's Inequality is often the right tool to prove concentration for these kinds of random variables. Below we state two versions of this inequality which are of special interest for us. Both were taken from [13], where the reader can also find their proofs.

Proposition 2.2 ((2.9) in Corollary 2.3 in [13]) *Suppose X is a sum of n independent random variables, each being 1 with probability p and 0 with probability $(1-p)$. If $0 < \varepsilon \leq 3/2$, then*

$$\Pr(|X-\mathbb{E}(X)| \geq \varepsilon\mathbb{E}(X)) \leq 2\mathrm{e}^{-\varepsilon^2\mathbb{E}(X)/3}.$$

Proposition 2.3 ((2.11) in Corollary 2.8 in [13]) *Suppose X is a sum of n independent random variables, each being either 1 or 0. If $x \geq 7\mathbb{E}(X)$, then*

$$\Pr(X \geq x) \leq \mathrm{e}^{-x}.$$

2.2.3 A martingale inequality

When proving that certain random variables are concentrated, we apply a martingale inequality which was developed in Section 3 in [2]. We now briefly explain the setup and the statement of this inequality.

Suppose that X is a random variable determined by m independent trials $T_1, \ldots, T_m$, each T_i being 1 with probability p_i, and 0 with probability $1-p_i$. Let c_i be the maximum change in the value of X when the outcome of T_i is altered. Set $C = \max c_i$.

Suppose also that, in order to obtain the outcome of trial T_i, we must pay $p_i(1-p_i)c_i^2$ dollars. Typically, the goal is to compute the value of X spending as little as possible. The way we evaluate X is by looking at outcomes of trials, one at a time, until we have enough information to compute X. We start by checking the outcome of a certain trial T_{i_1} of our choice. Based on the outcome of T_{i_1}, we choose an index i_2, and look at the outcome of T_{i_2}. Again, we choose i_3 and look at the outcome of T_{i_3}, and so on. In general, each time we choose the trial we want to observe based on the outcomes of trials previously observed. A *strategy* to compute X can be thought of as rooted tree that indicates which trials should be observed. Each node corresponds to a trial, and each node is either a leaf or branches into two other nodes, one for each possible outcome of its associated trial.

A strategy to compute X is said to have cost at most a, if no matter what the outcomes of $T_1, \ldots, T_m$ are, someone looking at the trials according to this strategy would always spend at most a dollars to obtain all necessary information to compute X.

Theorem 2.4 (Alon, Kim, Spencer [2]) *Suppose there exists a strategy to compute X of cost at most σ^2. If $\alpha C < 2\sigma$, then*

$$\Pr(|X - \mathbb{E}(X)| \geq \alpha\sigma) \leq 2\mathrm{e}^{-\alpha^2/4}.$$

See [3] page 101, for a proof of a slightly more general version of Theorem 2.4.

Chapter 3

A simple proof

3.1 Introduction

The goal of this chapter is to give a new proof to Conjecture 1.5 first stated by P. Erdős and H. Hanani in 1963. Until now, all known proofs of this conjecture use the "semi-random method", an iterative process in which the desired (n, k, ℓ)-system is built as a successive union of small randomly chosen pieces. The proof described in this chapter is based on a somewhat different approach. While our proof gives a much weaker bound on the proportion of uncovered ℓ-subsets than the one in [22], it has the advantage of being elementary and fairly simple.

We remind the reader of some basic definitions. An (n, k, ℓ)-system is a family $\mathscr{P}$ of k-element subsets of $[n]$ with the property that every ℓ-subset $L \subseteq [n]$ is contained in at most one member of $\mathscr{P}$. The maximum size of an (n, k, ℓ)-system is denoted by $S(n, k, \ell)$. Clearly,

$$S(n,k,\ell) \le \tbinom{n}{\ell}\tbinom{k}{\ell}^{-1}.$$

Erdős and Hanani [8] conjectured that

$$\lim_{n\to\infty} S(n,k,\ell)\tbinom{n}{\ell}^{-1}\tbinom{k}{\ell} = 1. \tag{3.1}$$

For an (n,k,ℓ)-system $\mathscr{P}$, let $\varepsilon(\mathscr{P})$ be the fraction of uncovered ℓ-subsets, i.e. $\varepsilon(\mathscr{P})$ satisfies

$$|\mathscr{P}| = (1-\varepsilon(\mathscr{P}))\binom{n}{\ell}\binom{k}{\ell}^{-1}.$$

Let $\varepsilon(n) = \varepsilon_{k,\ell}(n)$ denote the minimum $\varepsilon(\mathscr{P})$ among all (n,k,ℓ)-systems $\mathscr{P}$, i.e. let $\varepsilon(n)$ satisfy

$$S(n,k,\ell) = (1-\varepsilon(n))\binom{n}{\ell}\binom{k}{\ell}^{-1}.$$

We can rephrase Erdős-Hanani conjecture as

$$\lim_{n\to\infty} \varepsilon(n) = 0. \tag{3.2}$$

3.2 An easy bound

Set $\rho_0 = \frac{1}{2}\binom{k}{\ell}^{-1}$. We prove the following proposition.

Proposition 3.1 *Any family $\mathscr{A}$ of k-sets contains a subfamily $\mathscr{B} \subseteq \mathscr{A}$ which is an (n,k,ℓ)-system of cardinality*

$$|\mathscr{B}| \geq |\mathscr{A}|\rho_0\binom{n-\ell}{k-\ell}^{-1}. \tag{3.3}$$

A consequence of Proposition 3.1 is the existence of an (n,k,ℓ)-system $\mathscr{Q}$ with $\varepsilon(\mathscr{Q}) \leq (1-\rho_0)$ as follows. Consider the family of all possible k-subsets of $[n]$. Then Proposition 3.1 gives an (n,k,ℓ)-system $\mathscr{Q} \subseteq \binom{[n]}{k}$ with

$$|\mathscr{Q}| \geq \rho_0\binom{n}{k}\binom{n-\ell}{k-\ell}^{-1}. \tag{3.4}$$

After simplifying the binomial coefficients (using Proposition 5.1 in the Appendix), we obtain

$$|\mathscr{Q}| \geq \rho_0\binom{n}{\ell}\binom{k}{\ell}^{-1}. \tag{3.5}$$

As desired, the size of $\mathscr{Q}$ is a ρ_0-fraction of the best possible size of an (n,k,ℓ)-system.

Proof of Proposition 3.1. The proof consists of a simple probabilistic argument. Set

$$p = \binom{k}{\ell}^{-1}\binom{n-\ell}{k-\ell}^{-1}.$$

We select k-subsets in $\mathscr{A}$ independently, each with probability p. Let $\tilde{\mathscr{B}}$ denote the resulting random set system. We call a pair of members of $\tilde{\mathscr{B}}$ *conflicting*, if they intersect in at least ℓ elements. Observe that, if we remove one member of $\tilde{\mathscr{B}}$ from each conflicting pair, we are left with an (n, k, ℓ)-system. The idea is to use the first moment method to show that, after such members are removed from $\tilde{\mathscr{B}}$, the resulting family still has large cardinality. We consider two random variables $X = X(\tilde{\mathscr{B}})$, and $Y = Y(\tilde{\mathscr{B}})$. The first is defined as $X = |\mathscr{B}|$, and Y counts the number of (unordered) conflicting pairs in $\tilde{\mathscr{B}}$. Since any $A \in \mathscr{A}$ can form a conflicting pair with at most $\binom{k}{\ell}\binom{n-\ell}{k-\ell}$ other members of $\mathscr{A}$, by the linearity of the expectation, we obtain

$$\begin{aligned}\mathbb{E}(X-Y) &= \mathbb{E}(X) - \mathbb{E}(Y)\\ &\geq |\mathscr{A}|p - \tfrac{1}{2}|\mathscr{A}|\binom{k}{\ell}\binom{n-\ell}{k-\ell}p^2\\ &= |\mathscr{A}|\tfrac{1}{2}\binom{k}{\ell}^{-1}\binom{n-\ell}{k-\ell}^{-1}\\ &= |\mathscr{A}|\rho_0\binom{n-\ell}{k-\ell}^{-1}.\end{aligned}$$

Consider a set system $\mathscr{B}_0 \subseteq \mathscr{A}$ with $X(\mathscr{B}_0) - Y(\mathscr{B}_0) \geq \mathbb{E}(X-Y)$. Removing from $\mathscr{B}_0$ one member for every conflicting pair (at most Y members), we obtain an (n, k, ℓ)-system $\mathscr{B}$ satisfying (3.3). □

3.3 The construction

In [16], N. Kuzjurin studied a variant of this problem in which k is not constant, but varies with n. He considered an algebraic construction (see also Zinoviev [26]) to prove that, if $k = k(n)$ tends to infinity, $\ell = \ell(n) = o(k(n))$, and $k(n) < c\sqrt{n}$ for some $c < 1$, then (3.1) is true. We use a similar construction to show (3.1) for fixed k and ℓ.

Lemma 3.2 *Let $n = qm$, where q is a prime power and $q \geq m$, and let $\mathscr{M}$ be an (m, k, ℓ)-system. Then there exists an (n, k, ℓ)-system $\mathscr{P}$ with*

$$\varepsilon(\mathscr{P}) \leq \varepsilon(\mathscr{M}) - \rho_0\, \varepsilon(\mathscr{M})^{\binom{k}{\ell}} + O\left(\tfrac{1}{m}\right), \tag{3.6}$$

where the constant in the big O term depends on k and ℓ only.

Proof. We identify the sets $[n]$ and $[m] \times [q]$, and partition $[n]$ into m sets $S_0, \ldots, S_{m-1}$ where $S_i = \{(i, j) \colon j \in [q]\}$. A subset T of $[n]$ is said to be *crossing* if $|T \cap S_i| \leq 1$ for every i, $0 \leq i < m$. Let $\mathscr{L}$ be the family of all polynomials of degree $\ell - 1$ over $\mathbb{F}_q$, the field on q elements. We associate every polynomial $f \in \mathscr{L}$ with a crossing m-subset of $[n]$ by making $T(f) = \{(i, j) \colon 0 \leq i < m, f(i) = j\}$. Let $\mathscr{T} = \{T(f) \colon f \in \mathscr{L}\}$. It is a well known fact (c.f. Lagrange's Interpolation Formula) that if K is a field, then for any given distinct elements $x_1, \ldots, x_\ell \in K$ and for any $y_1, \ldots, y_\ell \in K$ there exists a unique polynomial f over K of degree $\ell - 1$ such that $f(x_i) = y_i$ for every i, $1 \leq i \leq \ell$. Therefore, every crossing ℓ-subset of $[n]$ determines precisely one set $T(f)$, in which it is contained. That implies $\mathscr{T}$ is an (n, m, ℓ)-system on $[n]$, and $|\mathscr{T}| = |\mathscr{L}| = q^\ell$.

Now, we will describe how to obtain an (n, k, ℓ)-system $\mathscr{R}$ from the (n, m, ℓ)-system $\mathscr{T}$ and the (m, k, ℓ)-system $\mathscr{M}$. For each m-subset $T(f) \in \mathscr{T}$, consider a copy of $\mathscr{M}$ with vertex set $T(f)$. The union of all such copies of $\mathscr{M}$ forms an (n, k, ℓ)-system $\mathscr{R}$. Observe that there are precisely $m!/|Aut(\mathscr{M})|$ distinct copies of $\mathscr{M}$ with vertex set $T(f)$. Consequently, there are

$$\left(\tfrac{m!}{|Aut(\mathscr{M})|}\right)^{|\mathscr{T}|}$$

distinct (n,k,ℓ)-systems obtained in the way described above. Let $\mathbb{R}$ be the set of all such (n,k,ℓ)-systems $\mathscr{R}$. Note that each $\mathscr{R}\in\mathbb{R}$ has cardinality

$$
\begin{aligned}
|\mathscr{R}| &= |\mathscr{T}||\mathscr{M}| \\
&= q^{\ell}(1-\varepsilon(\mathscr{M}))\tbinom{m}{\ell}\tbinom{k}{\ell}^{-1} \\
&= (1-\varepsilon(\mathscr{M}))\left(\tfrac{q(m-1)}{qm-1}\right)\left(\tfrac{q(m-2)}{qm-2}\right)\cdots\left(\tfrac{q(m-\ell+1)}{qm-\ell+1}\right)\tbinom{n}{\ell}\tbinom{k}{\ell}^{-1}. \qquad (3.7)
\end{aligned}
$$

For each $i=1,\dots,\ell-1$, we have $\frac{q(m-i)}{qm-i}\geq\frac{q(m-\ell)}{qm}=1-\frac{\ell}{m}$. Using that $(1-x)^{\ell}\geq(1-x\ell)$ for $0<x<1$, which one can easily prove by induction, we obtain

$$\prod_{i=1}^{\ell-1}\frac{q(m-i)}{qm-i}\geq 1-\frac{\ell^2}{m}. \qquad (3.8)$$

Combining (3.7) and (3.8) we conclude

$$|\mathscr{R}|\geq(1-\varepsilon(\mathscr{M}))\left(1-O\left(\tfrac{1}{m}\right)\right)\tbinom{n}{\ell}\tbinom{k}{\ell}^{-1} \qquad \forall\mathscr{R}\in\mathbb{R}. \qquad (3.9)$$

Since for any $\mathscr{R}\in\mathbb{R}$, the proportion of uncovered ℓ-subsets of $[n]$ is still too large for our purpose, we are going to select $\mathscr{R}_0\in\mathbb{R}$ so that one can add more k-subsets to obtain a sufficiently larger (n,k,ℓ)-system.

A k-subset of $[n]$ is said to be *diverse* if it is crossing, and all its ℓ-subsets belong to distinct members of $\mathscr{T}$. In order to estimate the number of diverse k-subsets, we note that to construct such a set, we can start with a crossing ℓ-subset $\{x_1,\dots,x_\ell\}$ and gradually add elements, choosing $x_{\ell+i}$ such that $\{x_1,\dots,x_{\ell+i}\}$ is crossing, and $x_{\ell+i}\notin T(f)$ for every f determined by an ℓ-subset of $\{x_1,\dots,x_{\ell+i-1}\}$. Thus, the number of diverse k-subsets of $[n]$ can be bounded by

$$
\begin{aligned}
\#\text{DIVERSE} &\geq \tbinom{m}{k}q^{\ell}(q-1)\left(q-\tbinom{\ell+1}{\ell}\right)\left(q-\tbinom{\ell+2}{\ell}\right)\cdots\left(q-\tbinom{k-1}{\ell}\right) \\
&= \tbinom{m}{k}q^{k}\left(1-O\left(\tfrac{1}{q}\right)\right) \qquad (3.10) \\
&= \tbinom{n}{k}\left(1-O\left(\tfrac{1}{m}\right)\right).
\end{aligned}
$$

Now, given $\mathscr{R} \in \mathbb{R}$, a crossing k-subset of $[n]$ is said to be *available* if it is diverse, and none of its ℓ-subsets are in $\mathscr{R}$. In order to obtain the desired $\mathscr{R}_0 \in \mathbb{R}$, consider a randomly and uniformly chosen $\mathscr{R} \in \mathbb{R}$. Observe that the probability that any fixed crossing ℓ-subset is uncovered is precisely $\varepsilon(\mathscr{M})$. Note also that, for every diverse k-subset K of $[n]$, all events of the form "L is not covered by $\mathscr{R}$", where L is an ℓ-subset of K, are mutually independent. Hence, the probability that a fixed diverse k-set K is available is precisely

$$\varepsilon(\mathscr{M})^{\binom{k}{\ell}}.$$

Denoting by $\textsc{available}(\mathscr{R})$ the set of available k-sets with respect to $\mathscr{R} \in \mathbb{R}$, we consequently have

$$\mathbb{E}(\#\textsc{available}(\mathscr{R})) = \#\text{DIVERSE} \times \varepsilon(\mathscr{M})^{\binom{k}{\ell}}. \tag{3.11}$$

On the other hand, fix $\mathscr{R}_0 \in \mathbb{R}$ such that $\textsc{available}(\mathscr{R}_0)$ has size

$$\#\textsc{available}(\mathscr{R}_0) \geq \mathbb{E}(\#\textsc{available}(\mathscr{R})). \tag{3.12}$$

Next, we add to $\mathscr{R}_0$ some available sets to obtain an (n, k, ℓ)-system $\mathscr{P}$ with the fraction $\varepsilon(\mathscr{P})$ of uncovered sets satisfying (3.6). Any single available k-set can be added to $\mathscr{R}_0$, but we cannot add too many of them at the same time, since we do not want the resulting family to have two members sharing ℓ points. In Section 3.2, we argued that any family $\mathscr{A}$ of k-sets contains an (n, k, ℓ)-system $\mathscr{B} \subseteq \mathscr{A}$ satisfying (3.3). Hence, one can fix an (n, k, ℓ)-system $\mathscr{B} \subseteq \textsc{available}(\mathscr{R}_0)$ with

$$|\mathscr{B}| \geq \#\textsc{available}(\mathscr{R}_0) \times \rho_0 \binom{n-\ell}{k-\ell}^{-1}. \tag{3.13}$$

We then construct the desired (n, k, ℓ)-system $\mathscr{P}$ as

$$\mathscr{P} = \mathscr{R}_0 \cup \mathscr{B}.$$

Now we need to estimate $|\mathscr{P}|$ in order to obtain the desired bound on $\varepsilon(\mathscr{P})$. We already have a lower bound on $|\mathscr{R}_0|$ given by inequality (3.9). It is left to bound the size of $\mathscr{B}$. From inequalities (3.11), (3.12), and (3.13) the following holds.

$$|\mathscr{B}| \geq \#\text{DIVERSE} \times \varepsilon(\mathscr{M})^{\binom{k}{\ell}} \times \rho_0 \binom{n-\ell}{k-\ell}^{-1}.$$

Using (3.10) and the binomial identity (5.1) (see Appendix, Proposition 5.1), we obtain

$$|\mathscr{B}| \geq \left(1 - O\left(\tfrac{1}{m}\right)\right) \varepsilon(\mathscr{M})^{\binom{k}{\ell}} \rho_0 \binom{n}{\ell} \binom{k}{\ell}^{-1}. \tag{3.14}$$

Finally, using (3.9) and (3.14), we are able to bound $|\mathscr{P}|$ from below:

$$|\mathscr{P}| \geq \left((1 - \varepsilon(\mathscr{M})) + \rho_0\, \varepsilon(\mathscr{M})^{\binom{k}{\ell}}\right)\left(1 - O\left(\tfrac{1}{m}\right)\right)\binom{n}{\ell}\binom{k}{\ell}^{-1}.$$

Therefore, by the definition of $\varepsilon(\mathscr{P})$, we have

$$(1 - \varepsilon(\mathscr{P})) \geq \left((1 - \varepsilon(\mathscr{M})) + \rho_0\, \varepsilon(\mathscr{M})^{\binom{k}{\ell}}\right)\left(1 - O\left(\tfrac{1}{m}\right)\right).$$

The desired inequality follows, and this finishes the proof of the lemma. □

3.4 The analysis

Theorem 3.3 *Let $\varepsilon(n)$ be as previously defined. Then*

$$\lim_{n\to\infty} \varepsilon(n) = 0.$$

Proof. Let us assume, for the purpose of contradiction, that the statement is false, and let $\varepsilon > 0$ be defined as

$$\varepsilon = \limsup_{n\to\infty} \varepsilon(n).$$

Let δ be defined as

$$\delta = \tfrac{\rho_0}{3}\left(\tfrac{\varepsilon}{2}\right)^{\binom{k}{\ell}}.$$

Let $n_1 < n_2 < \dots$ be a sequence of integers satisfying $\varepsilon(n_i) \geq \varepsilon - \delta$ for every i. We want to apply Lemma 3.2 when $n = n_i$. Unfortunately, in order to apply Lemma 3.2, we need n of the form qm, where q is a prime power and $q \geq m$. To overcome this problem, we will apply Lemma 3.2 to a number $q_i m_i$ which is slightly larger than n_i (and an (m_i, k, ℓ)-system $\mathscr{M}_i$ which is optimal in the sense that $\varepsilon(\mathscr{M}_i) = \varepsilon(m_i)$). Lemma 3.2 then gives a $(q_i m_i, k, \ell)$-system $\mathscr{Q}_i$ on $[q_i m_i]$. We will show that the (n_i, k, ℓ)-system $\mathscr{P}_i$ induced by $\mathscr{Q}_i$ on $[n_i] \subseteq [q_i m_i]$ satisfies $\varepsilon(\mathscr{P}_i) < \varepsilon - \delta$ for i is sufficiently large, which is a contradiction to the assumption that $\varepsilon(n_i) \geq \varepsilon - \delta$ for all i.

Let $\alpha < 1$ be a number such that for any sufficiently large integer s there is a prime on the interval $[s, s + s^\alpha]$. (By a result of D. R. Heath-Brown and H. Iwaniec [12], any $\alpha > \frac{11}{20}$ satisfies this requirement.) For every $i = 1, 2, \dots,$ let $m_i = \lceil \sqrt{n_i}\, \rceil$, and let q_i be a prime in $[m_i, m_i + m_i^\alpha]$. Also, consider an (m_i, k, ℓ)-system $\mathscr{M}_i$ with $\varepsilon(\mathscr{M}_i) = \varepsilon(m_i)$. By Lemma 3.2, there exists an $(q_i m_i, k, \ell)$-system $\mathscr{Q}_i$ satisfying

$$\varepsilon(\mathscr{Q}_i) \leq \varepsilon(m_i) - \rho_0\, \varepsilon(m_i)^{\binom{k}{\ell}} + O\big(n_i^{-\frac{1}{2}}\big).$$

We now restrict $\mathscr{Q}_i$ to a system $\mathscr{P}_i$ on n_i points, by deleting a total of $q_i m_i - n_i = O(m_i^\alpha \sqrt{n_i}) = O\big(n_i^{\frac{1+\alpha}{2}}\big)$ points. The number of k-sets which are in $\mathscr{Q}_i$ but not in $\mathscr{P}_i$ is bounded by

$$O\big(n_i^{\frac{1+\alpha}{2}}\big)\binom{q_i m_i - 1}{\ell - 1}\binom{k-1}{\ell-1}^{-1} = O\big((q_i m_i)^{\ell - 1 + \frac{1+\alpha}{2}}\big).$$

Hence, the fraction of uncovered ℓ-subsets in $\mathscr{P}_i$ is given by

$$\begin{aligned} \varepsilon(\mathscr{P}_i) &\leq \varepsilon(\mathscr{Q}_i) + O\big((q_i m_i)^{-\frac{1-\alpha}{2}}\big) \\ &\leq \varepsilon(m_i) - \rho_0\, \varepsilon(m_i)^{\binom{k}{\ell}} + O\big(n_i^{-\frac{1}{2}}\big) + O\big(n_i^{-\frac{1-\alpha}{2}}\big). \end{aligned} \tag{3.15}$$

Since the constants in the big O terms depend on k and ℓ only, for sufficiently large i (and consequently also n_i), the last two terms are each strictly less than $\frac{\delta}{2}$. Therefore

$$\varepsilon(\mathscr{P}_i) < \varepsilon(m_i) - \rho_0\, \varepsilon(m_i)^{\binom{k}{\ell}} + \delta. \tag{3.16}$$

We will show that the right-hand-side of (3.16) is less than $\varepsilon - \delta$ which is a contradiction to the assumption on the sequence $(n_i)_{i=1}^{\infty}$ that $\varepsilon(n_i) \geq \varepsilon - \delta$ for every i.

If $\varepsilon(m_i) \leq \varepsilon - 2\delta$, inequality (3.16) clearly implies $\varepsilon(\mathscr{P}_i) < \varepsilon - \delta$. Otherwise, suppose $\varepsilon(m_i) > \varepsilon - 2\delta$. By the definition of ε, for i sufficiently large, $\varepsilon(m_i) \leq \varepsilon + \delta$ holds. Combining (3.16) with the definition of δ we obtain

$$\begin{aligned}\varepsilon(\mathscr{P}_i) &< (\varepsilon + \delta) - \rho_0\,(\varepsilon - 2\delta)^{\binom{k}{\ell}} + \delta \\ &\leq \varepsilon - \rho_0\left(\tfrac{\varepsilon}{2}\right)^{\binom{k}{\ell}} + 2\delta \\ &= \varepsilon - \delta,\end{aligned}$$

which completes the proof. □

3.5 A finer analysis

We remark that a more careful analysis of the construction in Lemma 3.2, yields $\varepsilon(n) \leq C(\log\log n)^{-\beta}$, where $\beta = 1/\left(\binom{k}{\ell} - 1\right)$. This will be accomplished by Lemma 3.4, Lemma 3.5 and Theorem 3.6 below.

Before we go any further, let $A = A(k,\ell)$ be a large enough constant so that the big O term in (3.6) can be bounded by A/m. Let t_0 be the smallest integer such that all numbers $t \geq t_0$ satisfy

$$\frac{\rho_0}{2} t^{-\beta\binom{k}{\ell}} \geq \frac{A}{2^{2^t}}. \tag{3.17}$$

Also, let

$$B \geq \max\left\{(1-\rho_0)t_0^{\beta}, \left(\tfrac{2\beta}{\rho_0}\right)^{\beta}\right\},$$

and set $C = B + 1$.

Lemma 3.4 *Let q be a prime power. Let m be an integer less than q and suppose $\varepsilon(m) \leq B(t-1)^{-\beta}$ for some integer $t > t_0$. Then*

$$\varepsilon(qm) \leq Bt^{-\beta}. \tag{3.18}$$

Proof. By Lemma 3.2, we know

$$\varepsilon(qm) \leq \varepsilon(m) - \rho_0 \varepsilon(m)^{\binom{k}{\ell}} + \frac{A}{m}. \tag{3.19}$$

Using simple methods from calculus, one can show that $x - ax^b$ is increasing on $(0,1)$ when $ab < 1$ and $b \geq 1$. Hence, we may replace $\varepsilon(m)$ by $B(t-1)^{-\beta}$ in (3.19) to obtain

$$\varepsilon(qm) \leq B(t-1)^{-\beta} - \rho_0 \big(B(t-1)^{-\beta}\big)^{\binom{k}{\ell}} + \frac{A}{m}.$$

Since $t - 1 \geq t_0$ satisfies (3.17), the previous inequality implies

$$\varepsilon(qm) \leq B(t-1)^{-\beta} - \frac{\rho_0}{2} \big(B(t-1)^{-\beta}\big)^{\binom{k}{\ell}}. \tag{3.20}$$

We want to show that the right-hand-side of (3.20) is bounded by $Bt^{-\beta}$. Equivalently, we need to show

$$\frac{1}{(t-1)^\beta} - \frac{1}{t^\beta} \leq \lambda \left(\frac{1}{(t-1)^\beta} \right)^{\binom{k}{\ell}}, \tag{3.21}$$

where $\lambda = \frac{\rho_0}{2} B^{\binom{k}{\ell}-1}$. To this end, consider the function $g(t) = 1/t^\beta$. By the Mean Value Theorem, we know $g(t-1) - g(t) = -g'(\theta)$ for some $\theta \in [t-1, t]$. Since $\beta\theta^{-\beta-1} \leq \beta(t-1)^{-\beta-1}$, it would suffice to prove

$$\frac{\beta}{(t-1)^{\beta+1}} \leq \frac{\lambda}{t^{\beta\binom{k}{\ell}}}.$$

Replacing $1/\big(\binom{k}{\ell} - 1\big)$ for β in the exponents, we can see that the previous inequality holds as long as

$$1 \leq \frac{\lambda}{\beta}. \tag{3.22}$$

Recalling that B is at least $\big(\frac{2\beta}{\rho_0}\big)^\beta$, one can check that (3.22) holds. The proof follows. $\square$

Lemma 3.5 *If $t \geq t_0$, then*

$$\varepsilon(2^{2^t}) \leq Bt^{-\beta}. \tag{3.23}$$

Proof. The proof is by induction on t. The base case is $t = t_0$. By Proposition 3.1 and the definition of B, we have

$$\varepsilon(2^{2^{t_0}}) \leq (1 - \rho_0) \leq Bt_0^{-\beta}.$$

Now suppose $t > t_0$, and assume (3.23) holds for $t - 1$. Let $q = m = 2^{2^{t-1}}$. By the induction hypothesis, we have $\varepsilon(m) \leq B(t-1)^{-\beta}$. Hence, the proof follows by Lemma 3.4. □

Theorem 3.6 *If n is sufficiently large, then*

$$\varepsilon(n) \leq C(\log\log n)^{-\beta}. \tag{3.24}$$

Proof. Let t be a number such that $2^{2^t} < n \leq 2^{2^{t+1}}$, and let $m = 2^{2^{t-1}}$. Note that $n^{1/4} \leq m < n^{1/2}$. We may assume n is large enough so that $t > t_0$. As in Theorem 3.3, let $\alpha < 1$ be a number such that for all sufficiently large s there is a prime in the interval $[s, s + s^\alpha]$. Set $s = n/m$. Let q be a prime in $[s, s + s^\alpha]$. By Lemma 3.4, we have

$$\varepsilon(qm) \leq B(t-1)^{-\beta} \leq B(\log\log n^{1/2})^{-\beta}. \tag{3.25}$$

Let $\mathscr{Q}$ be an optimal (qm, k, ℓ)-system, i.e. with $\varepsilon(\mathscr{Q}) = \varepsilon(qm)$. We remove all members of $\mathscr{Q}$ that contain elements from $\{qm, qm-1, \ldots, n+1\}$. Let $\mathscr{P}$ denote the resulting family. In particular, if $qm = n$, no member is removed, and we have $\mathscr{P} = \mathscr{Q}$. Observe that $\mathscr{P}$ can be viewed as an (n, k, ℓ)-system since its ground set is $[n]$. In summary, $\mathscr{P}$ is the restriction of $\mathscr{Q}$ to $[n]$. Therefore, we may upper bound $\varepsilon(\mathscr{P})$ as follows.

$$\varepsilon(\mathscr{P}) \leq \varepsilon(\mathscr{Q}) + \frac{|\mathscr{Q} \setminus \mathscr{P}|}{\binom{n}{\ell}/\binom{k}{\ell}}. \tag{3.26}$$

We want to find a bound on the rightmost term of (3.26). First, notice every element in $[qm]$ is contained in at most $\binom{qm-1}{\ell-1}$ ℓ-subsets of $[qm]$. Each such ℓ-subset, in turn, is contained in at most one member of $\mathscr{Q}$, which implies

$$|\mathscr{Q} \setminus \mathscr{P}| \leq (qm-n)\binom{qm-1}{\ell-1}. \tag{3.27}$$

Note also that $qm-n$ can be bounded as follows.

$$qm - n \leq (s+s^{\alpha})m - n \leq n^{\alpha}m^{1-\alpha} \leq n^{\frac{1+\alpha}{2}}. \tag{3.28}$$

We now expand $\binom{qm-1}{\ell-1}/\binom{n}{\ell}$ as

$$\left(\frac{qm-1}{\ell-1} \times \frac{qm-2}{\ell-2} \times \cdots \times \frac{qm-\ell+1}{1}\right)\left(\frac{n}{\ell} \times \frac{n-1}{\ell-1} \times \cdots \times \frac{n-\ell+1}{1}\right)^{-1}$$

and cancel repeated factors to get

$$\begin{aligned}
\binom{qm-1}{\ell-1}\binom{n}{\ell}^{-1} &= \frac{\ell}{n}\prod_{i=1}^{\ell-1}\frac{qm-i}{n-i} \\
&= \frac{\ell}{n}\prod_{i=1}^{\ell-1}\left(1+\frac{qm-n}{n-i}\right) \\
&\leq \frac{\ell}{n}\left(1+2n^{\frac{\alpha-1}{2}}\right)^{\ell}.
\end{aligned}$$

For n sufficiently large, we obtain

$$\binom{qm-1}{\ell-1}\binom{n}{\ell}^{-1} = 2^{\ell}\ell n^{-1}. \tag{3.29}$$

Finally, using (3.27) together with (3.28) and (3.29), one can upper bound the rightmost term of (3.26) by

$$2^{\ell}\ell\tbinom{k}{\ell}n^{\frac{\alpha-1}{2}}. \tag{3.30}$$

Combining the fact that $\mathscr{Q}$ is optimal with (3.25), (3.26), and (3.30) above yields

$$\varepsilon(n) \leq \varepsilon(\mathscr{P}) \leq \varepsilon(qm) + 2^{\ell}\ell\tbinom{k}{\ell}n^{\frac{\alpha-1}{2}} \leq B(\log\log n)^{-\beta} + (\log\log n)^{-\beta}.$$

Recall the definition of C on page 29, which gives $\varepsilon(n) \leq C(\log\log n)^{-\beta}$. $\square$

Chapter 4

Locally quasi-perfect packings

4.1 Introduction

In [4] N. Alon and R. Yuster proved as a consequence of their main result that for every two integers $1 < \ell < k$, and n tending to infinity, there exists an (n, k, ℓ)-system with the property that the number of uncovered ℓ-subsets containing each fixed vertex is $o(n^{\ell-1})$. Their proof, however, does not give any bounds on the proportion of uncovered ℓ-subsets. In this chapter, we show the existence of an (n, k, ℓ)-system such that the proportion of uncovered ℓ-subsets containing each j-subset, $0 \leq j < \ell$, is $O(n^{\ell-j-\beta} \ln^{\gamma} n)$ (see Theorem 1.11). This is a generalization of their result.

4.2 Simplifying the problem

To prove Theorem 1.11, it is enough to take care of subsets of size $\ell - 1$. That is, it is enough to show the following.

Theorem 4.1 *Let $\ell < k$ be fixed. For every n there exists an (n, k, ℓ)-system such that, for every $(\ell-1)$-subset $M \subseteq [n]$, the number of uncovered ℓ-subsets containing M is of order*

$$O(n^{1-\beta} \ln^{\gamma} n),$$

where $\beta = 1/\left(\binom{k}{\ell} - 1\right)$, *and* $\gamma > 0$ *is a constant depending on* k *and* ℓ.

Proposition 4.2 *Theorem 4.1 is equivalent to Theorem 1.11.*

Proof. Trivially, Theorem 1.11 implies Theorem 4.1. To prove the converse, we show that for every $0 \leq j < \ell$, every j-subset of $[n]$ is contained in at most $An^{\ell-j-\beta} \ln^\gamma n$ uncovered ℓ-subsets of $[n]$, for some constant A depending on k and ℓ. Theorem 4.1 settles the case $j = \ell - 1$, and defines the constant A. Assume now that $0 \leq j < \ell - 1$. Let J be an arbitrary j-subset of $[n]$. There are $\binom{n}{\ell-1-j}$ subsets of size $(\ell - 1)$ containing J, each of which is contained in at most $An^{1-\beta} \ln^\gamma n$ uncovered ℓ-subsets of $[n]$ (by Theorem 4.1). Therefore, the number of uncovered ℓ-subsets containing J is at most

$$A\binom{n}{\ell-1-j} n^{1-\beta} \ln^\gamma n \leq An^{\ell-j-\beta} \ln^\gamma n.$$

Theorem 1.11 follows. □

4.3 Main strategy

The basic structure of the proof is roughly that of Lemma 1 in [21]. We even try to use a similar notation, however the details are more involving due to the more general nature of our problem.

We now describe how we produce the desired (n, k, ℓ)-system $\mathcal{F}$. The main idea is to grow $\mathcal{F}$ iteratively. Initially, $\mathcal{F}$ is empty, and we have a family $\mathcal{H}$ of eligible k-sets. Among other things, we require $\mathcal{H}$ to be nearly regular, in the sense that every ℓ-subset of $[n]$ must be contained in approximately the same number of members of $\mathcal{H}$. In each iteration, we perform a "bite". A bite consists of selecting a few members of $\mathcal{H}$, adding them to $\mathcal{F}$, and then removing from $\mathcal{H}$ the selected k-sets together with those k-sets that intersect a selected one in at least ℓ elements. Furthermore, each iteration is conducted so that the resulting hypergraph $\mathcal{H}$ is nearly regular with respect to ℓ-sets (i.e. every ℓ-subset of vertices is contained in approximately the same number of members of $\mathcal{H}$). After a certain number of iterations, the number of eligible k-sets becomes so small that we are unable to guarantee near regularity. At this point we stop our procedure. The union of the selected k-subsets in the various iterations constitute the family $\mathcal{F}$. By construction, every ℓ-subset of $[n]$ belongs to at most one member of $\mathcal{F}$. One can show that $\mathcal{F}$ leaves very few ℓ-subsets uncovered.

This rough sketch does not explain how, in the end, every $(\ell - 1)$-set is contained in few uncovered ℓ-subsets. For this reason, after every iteration, in addition to near regularity, we also require every $(\ell - 1)$-set to be contained in approximately the same number of uncovered ℓ-subsets. Since very few ℓ-subsets are left uncovered, every $(\ell - 1)$-set must be contained in few uncovered ℓ-subsets.

Initially, we could take $\mathcal{H}$ to be the family of all k-subsets of $[n]$. However, that does not work for some technical reason. To get around this problem, we carefully choose a family $\mathcal{H}_0$ of eligible k-sets. At each step i, for $i = 1, 2, \ldots,$ the family of selected k-sets is called $\mathcal{S}_i$, and the family of the remaining eligible k-sets is called $\mathcal{H}_i$. If we are able to perform s steps, then $\mathcal{F} = \mathcal{S}_1 \cup \mathcal{S}_2 \cup \cdots \cup \mathcal{S}_s$, and $\mathcal{H}_0 \supseteq \mathcal{H}_1 \supseteq \cdots \supseteq \mathcal{H}_s$.

4.4 Preliminaries

4.4.1 Important definitions

A *hypergraph* is a pair $(V, \mathcal{H})$ where V is a set, and $\mathcal{H}$ is a family of subsets of V. The elements of V are called vertices and the elements of $\mathcal{H}$ are called hyperedges. We say that $(V, \mathcal{H})$ is *k-uniform* if $|E| = k$ for every $E \in \mathcal{H}$. Sometimes we refer to a hypergraph $(V, \mathcal{H})$ simply as $\mathcal{H}$.

Let V be a set, $\mathcal{F}$ a family of subsets of V, and i a positive integer. We denote by $\Delta_i(\mathcal{F})$ the family of all subsets of V of size i *contained* in some member of $\mathcal{F}$. Similarly, $\nabla_i(\mathcal{F})$ denotes the family of all subsets of V of size i *containing* some member of $\mathcal{F}$. If F is simply a subset of V, we write $\Delta_i(F)$ and $\nabla_i(F)$ for $\Delta_i(\{F\})$ and $\nabla_i(\{F\})$ respectively.

For $X, Y \subseteq [n]$, with $|X| < \ell$ and $|Y| < k$, we define $\deg_\ell(X)$ and $\deg(Y)$ as follows:

$$\deg_\ell(X) = |\{L \in \Delta_\ell(\mathcal{H})\colon L \supset X\}|,$$
$$\deg(Y) = |\{H \in \mathcal{H}\colon H \supset Y\}|.$$

Since the hypergraph $\mathcal{H}$ is k-uniform, we unify the notation and write $\deg_k(Y)$ for $\deg(Y)$. Below, we introduce the two main objects which will be used in the proofs.

Definition 4.3 *Given positive integers* $1 < \ell < k < n, d, D$, *and a function* $f\colon \mathbb{R} \to \mathbb{R}^+$, *an* (f, ℓ, k, n, d, D)*-graph, is a* k*-uniform hypergraph* $(V, \mathcal{H})$ *on* n *vertices satisfying*

(i) $$D - f(D) \le \deg_k(L) \le D \qquad \forall\, L \in \Delta_\ell(\mathcal{H}),$$

(ii) $$d - f(d) \le \deg_\ell(M) \le d \qquad \forall\, M \in \binom{V}{\ell-1},$$

(iii) $$\deg_k(I) \le (\ell + 7)\log n \qquad \forall\, I \in \binom{V}{\ell+1}.$$

Until f *is precisely defined in* (4.9) *in Section 4.5, we implicitly assume that* $f(x) = o(x)$.

In this chapter, edges of $\mathcal{H}$ are called k*-blocks*. Members of $\Delta_\ell(\mathcal{H})$, which are "shadows" of k-blocks, are called ℓ*-shadows*.

Definition 4.4 *A bite from a* k*-uniform hypergraph* $(V, \mathcal{H})$ *is an ordered pair* $(\mathcal{K}, \mathscr{W})$, *where* $\mathcal{K} \subseteq \mathcal{H}$, *and* $\mathscr{W} \subseteq \Delta_\ell(\mathcal{H})$. *A* k*-block in* $\mathcal{K}$ *which does* *not share ℓ elements with any other member of $\mathcal{K}$ is called selected. The ℓ-shadows in $\mathscr{W}$ are called wasted. The family of selected k-blocks is denoted*

$$\mathcal{S} = \{K \in \mathcal{K} \colon |K \cap K'| < \ell \text{ for all } K' \in \mathcal{K}, K' \neq K\}.$$

Given a bite $(\mathcal{K}, \mathscr{W})$ from $(V, \mathcal{H})$, we define the remaining hypergraph $(V, \mathcal{H}^*)$ as the set of k-blocks in $\mathcal{H}$ which are not selected, do not share ℓ elements with a selected k-block, and do not contain a wasted ℓ-shadow. That is

$$\mathcal{H}^* = \mathcal{H} \setminus \big(\mathcal{S} \cup \nabla_k(\Delta_\ell(\mathcal{S})) \cup \nabla_k(\mathscr{W})\big). \tag{4.1}$$

Note that, the definition of $\mathcal{H}^*$ yields

$$\Delta_\ell(\mathcal{H}^*) \subseteq \Delta_\ell(\mathcal{H}) \setminus (\Delta_\ell(\mathcal{S}) \cup \mathscr{W}). \tag{4.2}$$

In other words, the ℓ-shadows of $\mathcal{H}$ which are wasted or contained in some selected k-block are no longer ℓ-shadows of $\mathcal{H}^*$. We denote the right hand side of (4.2) *by $\mathcal{L}^*$.*

4.4.2 Random bites

Given an (f, ℓ, k, n, d, D)-graph $(V, \mathcal{H})$, we define a random bite $(\mathcal{K}, \mathscr{W})$ in the following way. Independently, for every k-block $K \in \mathcal{H}$ and for every ℓ-shadow $L \in \Delta_\ell(\mathcal{H})$, we set

$$\Pr(K \in \mathcal{K}) = 1/D \quad \text{and} \quad \Pr(L \in \mathscr{W}) = w(L),$$

where $w(L)$ will be defined as follows. Let $p(L)$ be the probability that L is contained in some selected k-block $K \in \mathcal{S}$, and define

$$p^* = \max_{L \in \Delta_\ell(\mathcal{H})} p(L).$$

Now define the probability that $L \in \mathscr{W}$ as

$$w(L) = \frac{p^* - p(L)}{1 - p(L)}.$$

We implicitly assume that $D \geq 2$, so that $p(L) < 1$, which is important in order to have $w(L)$ well defined. The purpose of having a set of wasted ℓ-shadows is to guarantee that all ℓ-shadows of $\mathcal{H}$ have the same probability of being in $\mathcal{L}^*$. Namely, for every $L \in \Delta_\ell(\mathcal{H})$, we have

$$\begin{aligned}\Pr(L \in \mathcal{L}^*) &= \Pr(L \notin \Delta_\ell(\mathcal{S}) \wedge L \notin \mathscr{W}) \\ &= \Pr(L \notin \Delta_\ell(\mathcal{S}))\Pr(L \notin \mathscr{W}) \\ &= (1 - p(L))(1 - w(L)) \\ &= 1 - p^*.\end{aligned}$$

Small Waste

Below, we estimate the amount of wasted ℓ-shadows in a single random bite. Given a k-block $K \in \mathcal{H}$, define

$$\mathcal{N}(K) = \mathcal{H} \cap \nabla_k(\Delta_\ell(K)) \setminus \{K\}. \tag{4.3}$$

Note that $K \in \mathcal{S}$ holds if and only if $K \in \mathcal{K}$ and none of the blocks in $\mathcal{N}(K)$ is in $\mathcal{K}$. Hence

$$\Pr(K \in \mathcal{S}) = \frac{1}{D}\left(1 - \frac{1}{D}\right)^{|\mathcal{N}(K)|}. \tag{4.4}$$

By the definition of $p(L)$, and by the fact that two k-blocks sharing ℓ points cannot be simultaneously in $\mathcal{S}$, we have

$$\begin{aligned}p(L) &= \Pr(L \in \Delta_\ell(\mathcal{S})) \\ &= \sum_{K \in \mathcal{H} \cap \nabla_k(L)} \Pr(K \in \mathcal{S}).\end{aligned} \tag{4.5}$$

Using (4.4), we obtain

$$p(L) = \sum_{K \in \mathcal{H} \cap \nabla_k(L)} \frac{1}{D}\left(1 - \frac{1}{D}\right)^{|\mathcal{N}(K)|}. \tag{4.6}$$

Observe that $|\mathcal{N}(K)| = \binom{k}{\ell}D + O(f(D))$ for any block K, hence

$$\begin{aligned} p(L) &= (D + O(f(D)))\frac{1}{D}\Big(1 - \frac{1}{D}\Big)^{\binom{k}{\ell}D+O(f(D))} \\ &= (1 + O(f(D)/D))e^{-\binom{k}{\ell}}. \end{aligned}$$

In particular, for the maximum of the $p(L)$, we also have

$$p^* = (1 + O(f(D)/D))e^{-\binom{k}{\ell}}. \tag{4.7}$$

Using the definition of $w(L)$, and recalling that $f(D) = o(D)$, we obtain

$$w(L) = O(f(D)/D), \tag{4.8}$$

for all $L \in \Delta_\ell(\mathcal{H})$.

4.5 Biting lemma

From now on, set

$$f(x) = Ax^{1/2}\ln^{3/2} n, \tag{4.9}$$

where A is a large constant depending on k and ℓ which is chosen to satisfy inequalities (4.19), (4.23), and (4.30) in the proof of Lemma 4.6.

Definition 4.5 *A bite $(\mathcal{K}, \mathscr{W})$ from an (f, ℓ, k, n, d, D)-graph $(V, \mathcal{H})$ is said to be a* good bite *if it satisfies the following.*

(a) For any fixed $(\ell - 1)$-subset $M \subseteq V$, the number of wasted ℓ-shadows containing M is at most $21AdD^{-1/2}\ln^{3/2} n$.

(b) The remaining hypergraph $(V, \mathcal{H}^*)$ is an $(f, \ell, k, n, d^*, D^*)$-graph, with d^* and D^* given by

$$d^* = d(1 - p^*) + Bd^{1/2}\ln^{3/2} n, \tag{4.10}$$

$$D^* = D(1 - p^*)^{\binom{k}{\ell}-1} + BD^{1/2}\ln^{3/2} n, \tag{4.11}$$

where B is a constant depending on k and ℓ, chosen to satisfy inequality (4.28) *in the proof of Lemma 4.6.*

(c) $\mathcal{L}^* = \Delta_\ell(\mathcal{H}^*)$

Lemma 4.6 *If $(V, \mathcal{H})$ is an (f, ℓ, k, n, d, D)-graph, then a random bite from $(V, \mathcal{H})$ is a.a.s. a good bite provided*

$$d \geq D > \left(4A\tbinom{k}{\ell}\right)^2 \ln^3 n. \tag{4.12}$$

Proof. We divide the proof into three main parts. In Subsection 4.5.1, we show that a.a.s. a random bite satisfies (a) in Definition 4.5.

Second, we have to check that a.a.s. $\mathcal{H}^*$ satisfies property (i) in Definition 4.3 with respect to D^*. For an ℓ-subset $L \subseteq V$, let $\deg_k^*(L)$ denote the number of k-blocks in $\mathcal{H}^*$ containing L. In Subsection 4.5.2, we show that a.a.s. all $L \in \mathcal{L}^*$ satisfy

$$D^* - f(D^*) \leq \deg_k^*(L) \leq D^*. \tag{4.13}$$

Moreover (4.13) implies $\mathcal{L}^* = \Delta_\ell(\mathcal{H}^*)$, which takes care of (c) in Definition 4.5.

Finally, the last part corresponds to verifying that a.a.s. $\mathcal{H}^*$ satisfies property (ii) in Definition 4.3 with respect to d^*. For an $(\ell - 1)$-subset $M \subseteq V$, let $\deg_\ell^*(M)$ denote the number of ℓ-shadows of $\mathcal{H}^*$ containing M. In Subsection 4.5.3, we show that a.a.s. all $(\ell - 1)$-subsets $M \subseteq V$ satisfy

$$d^* - f(d^*) \leq \deg_\ell^*(M) \leq d^*. \tag{4.14}$$

Note also that, since $\mathcal{H}^*$ is a sub-hypergraph of $\mathcal{H}$, it automatically satisfies property (iii) in Definition 4.3. Hence, the second and third parts of the proof guarantees that a.a.s. a random bite satisfies (b) in Definition 4.5.

4.5.1 Bounding the wasted ℓ-degree of $(\ell-1)$-sets

We prove that, with probability tending to 1 with n, a random bite satisfies part (a) of Definition 4.5. Together, Claim 4.7 and Claim 4.8 will accomplish this task.

Let M be an arbitrary $(\ell-1)$-subset of vertices, and let Z_M be the random variable counting the number of wasted ℓ-shadows containing M.

Claim 4.7 $\mathbb{E}(Z_M) \leq 3AdD^{-1/2}\ln^{3/2} n$.

Proof. First, we need to estimate $p(L)$ as well as $w(L)$ more precisely. Recall, by equation (4.6), that

$$p(L) = \sum_{K\in\mathcal{H}\cap\nabla_k(L)} \frac{1}{D}\left(1-\frac{1}{D}\right)^{|\mathcal{N}(K)|}.$$

To find bounds for $p(L)$, we combine (ii) in Definition 4.3 with the definition of $|\mathcal{N}(K)|$. We start with a lower bound on $p(L)$ for any $L \in \Delta_\ell(\mathcal{H})$. Since, $|\mathcal{N}(K)| \leq \binom{k}{\ell}(D-1)$, we have

$$p(L) \geq \left(1-\frac{f(D)}{D}\right)\left(1-\frac{1}{D}\right)^{\binom{k}{\ell}(D-1)}.$$

From the definition of f and Proposition 5.3 in the Appendix, it follows that

$$p(L) \geq \mathrm{e}^{-\binom{k}{\ell}}\big(1-AD^{-1/2}\ln^{3/2} n\big). \tag{4.15}$$

On the other hand, the corresponding lower bound on $|\mathcal{N}(K)|$ implies

$$\begin{aligned} p(L) &\leq \left(1-\frac{1}{D}\right)^{\binom{k}{\ell}(D-1-AD^{1/2}\ln^{3/2} n)} \\ &\leq \mathrm{e}^{-\binom{k}{\ell}(1-2AD^{-1/2}\ln^{3/2} n)}. \end{aligned}$$

Using the assumption on D, and the fact that $\mathrm{e}^x \leq 1+2x$ for $0 \leq x < \frac{1}{2}$, we have

$$p(L) \leq \mathrm{e}^{-\binom{k}{\ell}}\Big(1+4\binom{k}{\ell}AD^{-1/2}\ln^{3/2} n\Big). \tag{4.16}$$

Now, in order to find an upper bound on $w(L)$, we do the following. By (4.16) and the assumption that $D > \left(4A\binom{k}{\ell}\right)^2 \ln^3 n$, we have

$$1 - p(L) \geq 1 - e^{-\binom{k}{\ell}}\left(1 + 4\binom{k}{\ell} A D^{-1/2} \ln^{3/2} n\right) \geq \tfrac{1}{2}. \tag{4.17}$$

Combining (4.15), (4.16), and (4.17) we obtain

$$\begin{aligned} w(L) &= \frac{p^* - p(L)}{1 - p(L)} \\ &\leq 2(1 + 4\tbinom{k}{\ell})\, e^{-\binom{k}{\ell}} A D^{-1/2} \ln^{3/2} n \\ &\leq 3AD^{-1/2} \ln^{3/2} n. \end{aligned} \tag{4.18}$$

Now, at most d ℓ-shadows contain the fixed $(\ell - 1)$-subset M. Each one of them is wasted with probability at most $3AD^{-1/2} \ln^{3/2} n$. By the linearity of expectation, we conclude

$$\mathbb{E}(Z_M) \leq 3AdD^{-1/2} \ln^{3/2} n.$$

The claim is proved. □

Let $\mathcal{E}_1(M)$ be the event that M is contained in at least $21AdD^{-1/2} \ln^{3/2} n$ wasted ℓ-shadows, and let $\mathcal{E}_1$ be the event that $\mathcal{E}_1(M)$ happens for some M.

Claim 4.8 $\Pr(\mathcal{E}_1) \to 0$ *as* $n \to \infty$.

Proof. Now we use Chernoff's Inequality to bound the probability of $\mathcal{E}_1(M)$. Using Claim 4.7, and Proposition 2.3, we obtain

$$\begin{aligned} \Pr(\mathcal{E}_1(M)) &= \Pr(Z_M \geq 21AdD^{-1/2} \ln^{3/2} n) \\ &\leq \exp\{-21AdD^{-1/2} \ln^{3/2} n\} \\ &\leq \exp\{-21A \ln n\}, \end{aligned}$$

where the last inequality follows from the assumption that $d \geq D > \ln^3 n$. So the probability that there exist $\ell - 1$ vertices that are contained in at least

$21AdD^{-1/2}\ln^{3/2} n$ wasted ℓ-shadows is

$$\Pr(\mathcal{E}_1) = \Pr\left(\bigcup \mathcal{E}_1(M)\right) \le \sum \Pr(\mathcal{E}_1(M)) \le n^{\ell-1} \times n^{-21A} \le n^{-1}. \quad (4.19)$$

Since $\Pr(\mathcal{E}_1)$ tends to zero as n tends to infinity, the claim follows. $\square$

4.5.2 Bounding the k-degree of ℓ-shadows

We need to show that a.a.s. (4.13) holds for every ℓ-shadow $L \in \mathcal{L}^*$. Here, certain difficulties arise. The first problem is that we do not know in advance which ℓ-shadows of $\mathcal{H}$ will still be in $\mathcal{L}^*$. So we need to prove such a statement for all ℓ-shadows of $\mathcal{H}$. Second, if L is not an ℓ-shadow of $\mathcal{H}^*$, then $\deg_k^*(L) = 0$, which implies that concentration does not hold for the random variable $\deg_k^*(L)$. To get around these problems, we consider a random variable X_L which is slightly different from $\deg_k^*(L)$. Let L be an arbitrary ℓ-shadow of $\mathcal{H}$. Define

$$\mathcal{R}(L) = \{K \in \mathcal{H} \colon \tbinom{K}{\ell} \setminus \{L\} \subseteq \mathcal{L}^*\}.$$

Note that $\mathcal{H}^* \subseteq \mathcal{R}(L)$. Let X_L be the number of k-blocks containing L in the hypergraph $(V, \mathcal{R}(L))$. In other words, X_L is the random variable defined by

$$X_L = |\nabla_k(L) \cap \mathcal{R}(L)|. \quad (4.20)$$

Note also that, whenever $L \in \Delta_\ell(\mathcal{H}^*)$, the family $\mathcal{R}(L)$ is exactly $\mathcal{H}^*$, and X_L is precisely $\deg_k^*(L)$, the number of k-blocks containing L in $(V, \mathcal{H}^*)$. We will show

$$D^* - f(D^*) \le X_L \le D^* \qquad \forall L \in \Delta_\ell(\mathcal{H}), \quad (4.21)$$

and that implies (4.13) as explained above.

We need two claims in order to prove (4.21). Their proofs will be given later in this subsection.

Claim 4.9 $\mathbb{E}(X_L) = \deg_k(L)\,(1-p^*)^{\binom{k}{\ell}-1}(1 + O(D^{-1}\ln n))$.

Before we state the second claim, let us define $\mathcal{E}_2(L)$ to be the event that

$$|X_L - \mathbb{E}(X_L)| > \tfrac{1}{2}BD^{1/2}\ln^{3/2} n,$$

where B, as mentioned earlier, is a large enough constant depending on k and ℓ, and chosen to satisfy inequality (4.28) in the proof of Claim 4.10. Let $\mathcal{E}_2$ be the event that $\mathcal{E}_2(L)$ happens for some $L \in \Delta_\ell(\mathcal{H})$.

Claim 4.10 $\Pr(\mathcal{E}_2) \to 0$ *as* $n \to \infty$.

Now, recall the definition of D^* in (4.11). By the previous claims, a.a.s. the following holds.

$$\begin{aligned} X_L &\le \deg_k(L)(1-p^*)^{\binom{k}{\ell}-1}(1+O(D^{-1})) + \tfrac{1}{2}BD^{1/2}\ln^{3/2} n \\ &\le D(1-p^*)^{\binom{k}{\ell}-1} + \tfrac{1}{2}BD^{1/2}\ln^{3/2} n + O(1). \end{aligned}$$

In particular, we have

$$X_L \le D(1-p^*)^{\binom{k}{\ell}-1} + BD^{1/2}\ln^{3/2} n = D^*. \tag{4.22}$$

The lower bound for $\deg_k^*(L)$ is obtained as follows.

$$\begin{aligned} X_L &\ge \big(D - AD^{1/2}\ln^{3/2} n\big)(1-p^*)^{\binom{k}{\ell}-1}(1+O(D^{-1})) - \tfrac{1}{2}BD^{1/2}\ln^{3/2} n \\ &= D(1-p^*)^{\binom{k}{\ell}-1} - \big(A(1-p^*)^{\binom{k}{\ell}-1} + \tfrac{1}{2}B\big)D^{1/2}\ln^{3/2} n + O(1) \\ &\ge D(1-p^*)^{\binom{k}{\ell}-1} - \big(A(1-p^*)^{(\binom{k}{\ell}-1)/2} - B\big)D^{1/2}\ln^{3/2} n \qquad (4.23) \\ &= D(1-p^*)^{\binom{k}{\ell}-1} + BD^{1/2}\ln^{3/2} n - A\big((1-p^*)^{\binom{k}{\ell}-1}D\big)^{1/2}\ln^{3/2} n \\ &\ge D^* - A(D^*)^{1/2}\ln^{3/2} n. \end{aligned}$$

This completes the proof of (4.21). We now proceed with the proofs of Claim 4.9 and Claim 4.10.

Proof of Claim 4.9. Fix $K \in \mathcal{H}$ containing L, let $t = \binom{k}{\ell} - 1$, and let $L_1, L_2, \ldots, L_t$ be the ℓ-shadows in $\binom{K}{\ell} \setminus \{L\}$. If we prove

$$\Pr\Big(\bigwedge_{i\in[t]} L_i \in \mathcal{L}^*\Big) = (1-p^*)^t(1+O(D^{-1}\ln n)), \tag{4.24}$$

then the claim follows by the linearity of expectation. The remainder of this proof is devoted to show (4.24).

Let $K_1, \dots, K_r \in \mathcal{H}$ be k-blocks, for some $1 \le r \le t$, no two of which share ℓ points. We first claim that

$$\Pr(K_1, \dots, K_r \in \mathcal{S}) = (1 + O(D^{-1} \ln n)) \prod_{i=1}^{r} \Pr(K_i \in \mathcal{S}). \tag{4.25}$$

Let $a = \Pr(K_1, \dots, K_r \in \mathcal{S})$, and let $b = \prod \Pr(K_i \in \mathcal{S})$. Let $\mathcal{N} = \bigcup \mathcal{N}(K_i)$. We can rewrite a and b as

$$a = \left(\prod_{i=1}^{r} \Pr(K_i \in \mathcal{K}) \right) \prod_{\tilde{K} \in \mathcal{N}} \Pr(\tilde{K} \notin \mathcal{K}),$$
$$b = \prod_{i=1}^{r} \left(\Pr(K_i \in \mathcal{K}) \prod_{\tilde{K} \in \mathcal{N}(K_i)} \Pr(\tilde{K} \notin \mathcal{K}) \right).$$

Note that a term of the form $\Pr(\tilde{K} \notin \mathcal{K})$ is counted multiple times in b but not in a whenever $\tilde{K}$ shares ℓ points with both K_i and K_j, for some $i \neq j$. Let $i \neq j$ be fixed, and let $\tilde{K}$ be such a block. Set $W = \tilde{K} \cap (K_i \cup K_j)$. Since W has at least $\ell + 1$ elements, there are only at most $(\ell + 7) \ln n$ possibilities for $\tilde{K}$. So, for a fixed choice of $i \neq j$, the number of blocks $\tilde{K}$ sharing ℓ elements with both K_i and K_j is bounded by $(\ell + 7)\binom{k}{\ell}^2 \ln n$. Hence at most $(\ell + 7)\binom{k}{\ell}^4 \ln n$ terms are missing in a, which implies (4.25) after a routine calculation.

Now let $F_1, \dots, F_r \in \binom{K}{\ell} \setminus \{L\}$, for some $1 \le r \le t$. Since two members of $\mathcal{H}$ sharing ℓ points cannot both be in $\mathcal{S}$, we have

$$\prod_{i=1}^{r} p(F_i) = \prod_{i=1}^{r} \sum^{\boldsymbol{A}} \Pr(\tilde{K} \in \mathcal{S}) = \sum^{\boldsymbol{B}} \prod_{i=1}^{r} \Pr(K_i \in \mathcal{S}), \tag{4.26}$$

where $\sum^{\boldsymbol{A}}$ is over all $\tilde{K} \in \mathcal{H}$ containing F_i, and $\sum^{\boldsymbol{B}}$ is over all r-tuples $(K_1, \dots, K_r)$ of k-blocks in $\mathcal{H}$ satisfying $F_i \subseteq K_i$ for every $i \in [r]$. We also know that

$$\begin{aligned}\Pr(F_1,\ldots,F_r \in \Delta_\ell(\mathcal{S})) &= \Pr(K \in \mathcal{S}) \\ &\quad + \sum^{\boldsymbol{C}} \Pr(K_1,\ldots,K_r \in \mathcal{S}) \\ &\quad + \sum^{\boldsymbol{D}} \Pr(K_1,\ldots,K_r \in \mathcal{S}),\end{aligned}$$

where $\sum^{\boldsymbol{C}}$ and $\sum^{\boldsymbol{D}}$ are both over all r-tuples $(K_1,\ldots,K_r)$ satisfying $F_i \subseteq K_i$ for every $i \in [r]$, the only difference being the following. In $\sum^{\boldsymbol{C}}$, one requires $|K_i \cap K_j| < \ell$ for every $1 \le i < j \le r$, while in $\sum^{\boldsymbol{D}}$ one requires that either $K_i = K_j$ or $|K_i \cap K_j| < \ell$ hold, with $K_i = K_j$ for at least one pair $i \neq j$. Notice that the terms in $\sum^{\boldsymbol{D}}$ contribute to $O(D^{-1}\ln n)$. Therefore $\Pr(F_1,\ldots,F_r \in \Delta_\ell(\mathcal{S}))$ can be written as

$$O(D^{-1}) + \sum^{\boldsymbol{C}} \Pr(K_1,\ldots,K_r \in \mathcal{S}) + O(D^{-1}\ln n).$$

Using (4.25), we obtain

$$O(D^{-1}\ln n) + (1 + O(D^{-1}\ln n)) \sum^{\boldsymbol{C}} \prod_{i=1}^{r} \Pr(K_i \in \mathcal{S}).$$

Recall that $\deg_k(F_i) = O(D)$, which implies $\sum^{\boldsymbol{C}}$ is missing $O(D^{r-1}\ln n)$ terms from the $\Theta(D^r)$ terms in $\sum^{\boldsymbol{B}}$ above. Therefore, we have

$$\Pr(F_1,\ldots,F_r \in \Delta_\ell(\mathcal{S})) = O(D^{-1}\ln n) + (1 + O(D^{-1}\ln n)) \sum^{\boldsymbol{B}} \prod_{i=1}^{r} \Pr(K_i \in \mathcal{S}),$$

which, combined with (4.26), yields

$$\Pr(F_1,\ldots,F_r \in \Delta_\ell(\mathcal{S})) = (1 + O(D^{-1}\ln n)) \prod_{i=1}^{r} p(F_i).$$

Since the wasted edges are chosen independently, if $G_1,\ldots,G_s \in \binom{K}{\ell} \setminus \{L\}$, for some $1 \le s \le t$, we also have that

$$\Pr\Big(\Big(\bigwedge_{i\in[r]} F_i \in \Delta_\ell(\mathcal{S})\Big) \wedge \Big(\bigwedge_{j\in[s]} G_j \in \mathscr{W}\Big)\Big)$$

equals

$$(1 + O(D^{-1} \ln n)) \prod_{i=1}^{r} p(F_i) \prod_{j=1}^{s} w(G_j). \tag{4.27}$$

Now recall $L_1, L_2, \ldots, L_t$ are the ℓ-shadows in $\binom{K}{\ell} \setminus \{L\}$, hence

$$\Pr\Big(\bigwedge_{i\in[t]} L_i \in \mathcal{L}^*\Big) = 1 - \Pr\Big(\Big(\bigvee_{i\in[t]} L_i \in \Delta_\ell(\mathcal{S})\Big) \vee \Big(\bigvee_{j\in[t]} L_j \in \mathscr{W}\Big)\Big).$$

By the Inclusion-Exclusion Formula (see Proposition 2.1 in Chapter 2), we can write $\Pr(\{L_1, L_2, \ldots, L_t\} \subseteq \mathcal{L}^*)$ as

$$\sum_{I,J\subseteq[t]} (-1)^{|I|+|J|} \Pr\Big(\Big(\bigwedge_{i\in I} L_i \in \Delta_\ell(\mathcal{S})\Big) \wedge \Big(\bigwedge_{j\in J} L_j \in \mathscr{W}\Big)\Big).$$

Using (4.27), we obtain

$$\sum_{I,J\subseteq[t]} (-1)^{|I|+|J|} \Big((1 + O(D^{-1} \ln n)) \prod_{i\in I} p(L_i) \prod_{j\in J} w(L_j)\Big).$$

Since the number of terms in the previous summation is constant (precisely 4^t), we have

$$\begin{aligned}\Pr\Big(\bigwedge_{i\in[t]} L_i \in \mathcal{L}^*\Big) &= (1 + O(D^{-1} \ln n)) \prod_{i\in[t]} (1 - p(L_i)) \prod_{j\in[t]} (1 - w(L_j)) \\ &= (1 - p^*)^t (1 - O(D^{-1} \ln n)).\end{aligned}$$

This settles (4.24), and Claim 4.9 follows. □

Proof of Claim 4.10. First, let $L \in \Delta_\ell(\mathcal{H})$ be fixed. We call a k-block *primary* if it contains L, and *secondary* if it is not primary, but shares ℓ points with some primary k-block. The number of primary k-blocks is no more than D, and the number of secondary k-blocks is less than $\binom{k}{\ell} D^2$. We call an ℓ-shadow *primary*, if it is not L and is contained in a primary k-block.

To prove that $\mathcal{E}_2(L) \to 0$ with n, we apply Theorem 2.4 to the random variable X_L, which clearly depends on the random bite $(\mathcal{K}, \mathscr{W})$. Note that a random bite can be viewed a sequence of trials, each of which is 1 if the corresponding k-block is in $\mathcal{K}$ (resp. ℓ-shadow is in $\mathscr{W}$), and 0 otherwise.

To apply the martingale bound, we first need to define a strategy to compute X_L consisting of a sequence of questions. Each question can be about a k-block being in $\mathcal{K}$ or about an ℓ-shadow being in $\mathscr{W}$. The price of a question of the first type is at most

$$\frac{1}{D}\left(1-\frac{1}{D}\right)c_k^2 < \frac{c_k^2}{D},$$

where c_k is the maximum possible change in the value of X_L when a k-block is added to $\mathcal{K}$ or removed from $\mathcal{K}$. Similarly, c_ℓ is the maximum possible change in the value of X_L when an ℓ-shadow is added to $\mathscr{W}$ or removed from $\mathscr{W}$. The price of a question about an ℓ-shadow F being in $\mathscr{W}$ is at most

$$w(F)\,(1-w(F))\,c_\ell^2 < c_\ell^2.$$

Estimating c_k and c_ℓ

Recall that

$$X_L = |\nabla_k(L) \cap \mathcal{R}(L)|,$$

and recall that $\mathcal{H}^*$ is defined in (4.1) as

$$\mathcal{H}^* = \mathcal{H} \setminus \big(\mathcal{S} \cup \nabla_k(\Delta_\ell(\mathcal{S})) \cup \nabla_k(\mathscr{W})\big).$$

Let us assume that a block K is removed from $\mathcal{K}$. The analysis regarding addition of a block into $\mathcal{K}$ is symmetric. The reasons why this operation could affect the value of X_L are two. First, because a block containing L could now belong to $\mathcal{S}$, although there can only be at most one such block. Second, because a few blocks containing L could be added to $\nabla_k(\Delta_\ell(\mathcal{S}))$. In fact, since blocks in $\mathcal{R}(L)$ are also counted in X_L, only the blocks added to $\nabla_k(\Delta_\ell(\mathcal{S}) \setminus \{L\})$ could possibly affect X_L.

To begin the analysis, note that $|\mathcal{S}|$ changes by at most $\binom{k}{\ell}$. This would happen precisely when there are $\binom{k}{\ell}$ blocks $\tilde{K}_0, \ldots, \tilde{K}_t$, each containing a distinct ℓ-shadow of K, and each of which was previously not in $\mathcal{S}$ only because K was in $\mathcal{K}$. In turn, for each block $\tilde{K}_i$ that goes into $\mathcal{S}$ (as a result of K being removed from $\mathcal{K}$), we will show that X_L changes by at most $k(\ell+7)\ln n$ as follows. If a block K' intersects $\tilde{K}_i$ in at least ℓ vertices, and the intersection is not L itself, then K' is not in $\mathcal{H}^* \cup \mathcal{R}(L)$. Since $\mathcal{H}$ is an (f, ℓ, k, n, d, D)-graph, every $\ell+1$ points are contained in at most $(\ell+7)\ln n$ blocks of $\mathcal{H}$ (see item (iii) in Definition 4.3). Hence every point in $\tilde{K}_i \setminus L$ is in at most $(\ell+7)\ln n$ blocks K' containing L. This settles the bound on the change of X_L with respect to $\tilde{K}_i$. Given the number of choices for $\tilde{K}_i$, we set $c_k = \binom{k}{\ell} k(\ell+7)\ln n$.

Suppose now that an ℓ-shadow $L' \neq L$ is removed from (or added to) $\mathscr{W}$. Since there are at most $(\ell+7)\ln n$ blocks of $\mathcal{H}$ containing both L' and L, we set $c_\ell = (\ell+7)\ln n$.

Defining the strategy to compute X_L

Let us define the sequence of questions that are necessary to compute X_L. First, we need to know whether each primary k-block is in $\mathcal{S}$. Hence, for each primary and secondary k-block K, we ask whether $K \in \mathcal{K}$. These questions are all of the first type, and their number is less than

$$\binom{k}{\ell} D^2.$$

At this point we already know which primary k-blocks are in $\mathcal{S}$, but we do not have enough information to determine which secondary k-blocks are in $\mathcal{S}$. We call a secondary k-block K a *candidate* to be in $\mathcal{S}$, if $K \in \mathcal{K}$, and none of the primary or secondary k-blocks which contain an ℓ-shadow of K are in $\mathcal{K}$. Deciding whether each candidate is in S requires less than $\binom{k}{\ell} D$ questions. Note that each candidate contains a primary ℓ-shadow, and each primary ℓ-shadow is contained in at most one candidate. Hence the number of such candidates is less than $\binom{k}{\ell}D$. Therefore, less than

$$\binom{k}{\ell}^2 D^2$$

additional questions of the first type are needed.

Finally, we need to know whether each primary ℓ-shadow F is in $\mathscr{W}$. The number of questions of the second type is less than

$$\binom{k}{\ell}D.$$

Now we have all the necessary information to determine how many k-blocks containing L survived in $\mathcal{H}^* \cup \mathcal{R}(L)$, which is precisely the value of X_L. The total cost of the algorithm can be estimated as

$$\binom{\#\text{ questions of}}{\text{the first type}} \frac{c_k^2}{D} + \binom{\#\text{ questions of}}{\text{the second type}} c_\ell^2 < BD\ln^2 n, \tag{4.28}$$

for some large enough constant B that depends on k and ℓ.

Applying the martingale inequality

Now we set $\alpha = 2\sqrt{(\ell+1)\ln n}$ and $\sigma^2 = BD\ln^2 n$, and apply Theorem 2.4 to the random variable X with parameters σ and α, obtaining

$$\Pr(|X_L - \mathbb{E}(X_L)| > 2\sqrt{(\ell+1)B}D^{1/2}\ln^{3/2} n) < 2e^{-\alpha^2/4} = 2n^{-(\ell+1)}. \tag{4.29}$$

We can bound $\Pr(\mathcal{E}_2(L))$ by the right hand side of (4.29), if we also assume that B is large enough to satisfy $B/2 \geq 2\sqrt{(\ell+1)B}$. Using the union bound, we can easily estimate $\Pr(\mathcal{E}_2)$ as follows.

$$\Pr(\mathcal{E}_2) = \Pr\left(\bigcup \mathcal{E}_2(L)\right) \leq \sum \Pr(\mathcal{E}_2(L)) \leq n^\ell \times 2n^{-(\ell+1)} \leq 2n^{-1}.$$

Clearly, $\Pr(\mathcal{E}_2) \to 0$ as n tends to infinity, and Claim 4.10 is proved. □

4.5.3 Bounding the ℓ-degree of $(\ell - 1)$-sets

Let M be an $(\ell-1)$-subset of V. Let Y_M be the random variable defined by $Y_M = \deg^*_\ell(M)$, the number of ℓ-shadows containing M in the hypergraph $(V, \mathcal{H}^*)$. The following claim easily follows from the linearity of expectation.

Claim 4.11 $\mathbb{E}(Y_M) = \deg_\ell(M)(1-p^*)$.

Now let $\mathcal{E}_3(M)$ be the event that

$$|Y_M - \mathbb{E}(Y_M)| > B' d^{1/2} \ln^{1/2} n$$

where B' is a constant depending on k and ℓ chosen to satisfy inequality (4.31) in the proof of Claim 4.12. Let $\mathcal{E}_3$ be the event that $\mathcal{E}_3(M)$ happens for some $(\ell-1)$-subset M. We are ready to state Claim 4.12.

Claim 4.12 $\Pr(\mathcal{E}_3) \to 0$ *as* $n \to \infty$.

We now use claims 4.11 and 4.12 to find upper and lower bounds for $\deg^*_\ell(M)$. The upper bound is easy, as a.a.s. we have

$$\deg^*_\ell(M) \le \deg_\ell(M)(1-p^*) + B' d^{1/2} \ln^{1/2} n \le d^*.$$

Regarding the lower bound, we know that a.a.s. $\deg^*_\ell(M)$ satisfies

$$\begin{aligned}
\deg^*_\ell(M) &\ge \deg_\ell(M)(1-p^*) - B' d^{1/2} \ln^{1/2} n \\
&\ge (d - A d^{1/2} \ln^{3/2} n)(1-p^*) - B' d^{1/2} \ln^{3/2} n \\
&= d(1-p^*) - \big(A(1-p^*) + B'\big) d^{1/2} \ln^{3/2} n \\
&\ge d(1-p^*) - \big(A(1-p^*)^{1/2} - B\big) d^{1/2} \ln^{3/2} n \qquad (4.30) \\
&= d(1-p^*) + B d^{1/2} \ln^{3/2} n - A\big((1-p^*)d\big)^{1/2} \ln^{3/2} n \\
&\ge d^* - A(d^*)^{1/2} \ln^{3/2} n.
\end{aligned}$$

Here, the dependency of A and B become clear, since we already mentioned that A needs to be large enough to satisfy (4.30).

Proof of Claim 4.12. This proof will basically follow along the lines of the proof of Claim 4.10. We fix an $(\ell-1)$-subset $M \subseteq V$. A k-block is *primary* if it contains M, and secondary if it is not primary and shares ℓ elements with some primary k-block.

We will apply Theorem 2.4 to bound the probability of $\mathcal{E}_3(M)$. As previously, a question about a k-block being in $\mathcal{K}$ costs $\frac{1}{D}\left(1-\frac{1}{D}\right)c_k^2 < c_k^2 D^{-1}$, and a question about an ℓ-shadow F being in $\mathscr{W}$ costs $w(F)(1-w(F))c_\ell^2 < c_\ell^2$. The only difference is that now we are dealing with a different random variable, so the values of c_k and c_ℓ need to be recomputed. That means, we need to find what is the maximum change in the value of Y_M when a k-block is added to (or removed from) $\mathcal{K}$, and also when an ℓ-shadow is added to (or removed from) $\mathscr{W}$.

This time, the analysis is a lot simpler. The only way an ℓ-shadow L containing M may be excluded from $\Delta_\ell(\mathcal{H}^*)$ (as a result of a k-block K being removed from $\mathcal{K}$) is if L is contained in some k-block $\tilde{K} \in \mathcal{K}$ which now belongs to $\mathcal{S}$. The maximum change in $|S|$ is still $\binom{k}{\ell}$, as before. Hence, the maximum change in Y_M is at most $c_k = (k-\ell+1)\binom{k}{\ell}$. On the other hand, it is easy to see that $c_\ell = 1$.

Now we define a strategy to compute the value of Y_M. Essentially, we need to know which ℓ-shadows containing M "survived the bite", i.e. which ℓ-shadows are neither in $\Delta_\ell(\mathcal{S})$ nor in $\mathscr{W}$. First, we ask whether each primary k-block is in $\mathcal{K}$. At most dD questions were asked. After that, some of the primary k-blocks could be candidates to be in $\mathcal{S}$. For each candidate $\tilde{K}$, we ask which secondary k-blocks (w.r.t. $\tilde{K}$) are in $\mathcal{K}$. Here, at most $\binom{k}{\ell}dD$ questions were asked. We now know exactly the primary k-blocks in $\mathcal{S}$, and consequently, the ℓ-shadows containing M that are in $\Delta_\ell(\mathcal{S})$.

Finally, to know which ℓ-shadows containing M were wasted, we need d additional questions. Therefore the total cost of the algorithm is at most

$$(dD + \tbinom{k}{\ell}dD)(k-\ell+1)\tbinom{k}{\ell}D^{-1} + d \leq B'd, \tag{4.31}$$

for some constant B' depending on k and ℓ.

Set $\rho = 2\sqrt{\ell \ln n}$, and $\sigma^2 = B'd$. We can now apply Theorem 2.4 with these parameters as follows.

$$\Pr\left(|Y_M - \mathbb{E}(Y_M)| > 2\sqrt{\ell B' d \ln n}\right) < 2e^{-\rho^2/4} = 2n^{-\ell}.$$

Assuming that B' is large enough to beat $2\sqrt{\ell B'}$, we can bound $\Pr(\mathcal{E}_3)$ by taking the union bound over all $(\ell-1)$-subsets of vertices. We obtain

$$\Pr(\mathcal{E}_3) = \Pr\left(\bigcup \mathcal{E}_3(M)\right) \leq \sum \Pr(\mathcal{E}_3(M)) \leq n^{\ell-1} \times 2n^{-\ell} \leq 2n^{-1}.$$

Claim 4.12 is proved. □

This completes the proof of Lemma 4.6. □

4.6 Construction of $\mathcal{H}_0$

The goal of this section is to prove the existence of an (f, ℓ, k, n, d, D)-graph $\mathcal{H}_0$ with $d = n$, $n \leq D \leq 2n \ln n$, and f given by (4.9).

If $\ell = k-1$, we simply set $\mathcal{H}_0 = \binom{[n]}{k}$. For $\ell < k-1$, consider the following random procedure. Choose each k-subset of $[n]$ independently with probability $p = n^{-k+\ell+1} \ln n$. Let $\mathcal{H}$ be the resulting random family. We are going to show that, with high probability, $\mathcal{H}$ is as desired. We then fix one such hypergraph to be $\mathcal{H}_0$.

For $L \in \binom{[n]}{\ell}$, consider the random variable $X_L = |\mathcal{H} \cap \nabla_k(L)|$. Similarly, for each $(\ell+1)$-subset of $[n]$, let $Y_I = |\mathcal{H} \cap \nabla_k(I)|$. Clearly, $\mathbb{E}(X_L) = n \ln n$, and $\mathbb{E}(Y_I) = \ln n$. Next, we argue that X_L and X_I are concentrated.

Starting with X_L, let us set $\epsilon = \sqrt{3(\ell+1)/\mathbb{E}(X_L)}$, and let A_L be the event that $|X_L - \mathbb{E}(X_L)| > \epsilon \mathbb{E}(X_L)$. By Chernoff's Inequality (Proposition 2.2), we have

$$\Pr(A_L) = \Pr\left(|X_L - \mathbb{E}(X_L)| > \epsilon \mathbb{E}(X_L)\right) < 2\exp\left\{-\tfrac{\epsilon^2}{3}\mathbb{E}(X_L)\right\} = 2n^{-\ell-1}.$$

Hence the probability that there exists $L \in \binom{[n]}{\ell}$ with $|X_L - \mathbb{E}(X_L)| > \sqrt{3(\ell+1)\, n \ln n}$ is less than $n^{\ell} \times 2n^{-\ell-1}$, which tends to 0 as n tends to infinity.

Now let $x = \max\{7, (\ell+2)\}\, \mathbb{E}(Y_I)$, and let B_I be the event that $Y_I > x$. Similarly, by (2.11) in Corollary 2.4 in [13], we can bound the upper tail of Y_I as

$$\Pr(B_I) = \Pr(Y_I \geq x) < e^{-x} \leq n^{-\ell-2}.$$

Therefore, the probability that there exists I with $Y_I > \max\{7, (\ell+2)\}\, \mathbb{E}(Y_I)$ is less than $n^{\ell+1} \times n^{-\ell-2}$, which tends to 0 as well.

In view of the previous discussion, we may fix a k-uniform hypergraph $\mathcal{H}_0$ on n vertices satisfying

$$|\nabla_\ell(M) \cap \Delta_\ell(\mathcal{H}_0)| = n - (\ell - 1) \qquad \forall M \in \binom{V}{\ell-1} \tag{4.32}$$

$$|\nabla_k(L) \cap \mathcal{H}_0| \geq n \ln n - (\ell+1)\sqrt{n \ln n} \qquad \forall L \in \binom{V}{\ell} \tag{4.33}$$

$$|\nabla_k(L) \cap \mathcal{H}_0| \leq n \ln n + (\ell+1)\sqrt{n \ln n} \qquad \forall L \in \binom{V}{\ell} \tag{4.34}$$

$$|\nabla_k(I) \cap \mathcal{H}_0| \leq (\ell+7) \ln n \qquad \forall I \in \binom{V}{\ell+1} \tag{4.35}$$

It is straightforward to check that (4.32) – (4.35) imply $\mathcal{H}_0$ is indeed an (f, ℓ, k, n, d, D)-graph with the desired parameters.

4.7 Successive bites

In order to prove Theorem 4.1, we study the evolution of the parameters d and D as we successively bite (with good bites) from $\mathcal{H}_0$. Set $d_0 = d$, and $D_0 = D$. We inductively define $\mathcal{H}_{i+1}$ to be the $(f, \ell, k, n, d_{i+1}, D_{i+1})$-graph that remains after a fixed good bite is taken from $\mathcal{H}_i$. By Lemma 4.6, there is always a good bite $(\mathcal{K}_{i+1}, \mathscr{W}_{i+1})$ from $\mathcal{H}_i$, provided D_i is not too small. Denote by $\mathcal{S}_{i+1}$ the set of selected k-blocks of the bite $(\mathcal{K}_{i+1}, \mathscr{W}_{i+1})$. By the definition of good bite, d_{i+1} and D_{i+1} must satisfy the relations given by (4.36) and (4.37) below.

$$d_{i+1} = d_i(1 - p_i^*) + Bd_i^{1/2}\ln^{3/2} n, \tag{4.36}$$

$$D_{i+1} = D_i(1 - p_i^*)^{\binom{k}{\ell}-1} + BD_i^{1/2}\ln^{3/2} n. \tag{4.37}$$

We denote by s the smallest integer such that $D_s > \ln^5 n$ is no longer satisfied. The parameter s denotes the number of bites that we perform. Next, we compute s, d_s and, for any fixed $(\ell - 1)$-subset $M \subset V$, we bound the number of ℓ-shadows containing M that were wasted along the s bites. Since $D_0 \leq 2d_0 \ln n$, relations (4.36) and (4.37) guarantee $D_i \leq 2d_i \ln n$ for all $0 \leq i \leq s$. Using this fact, and inequality (4.7), the relationship between d_{i+1} and d_i can be written as follows.

$$\begin{aligned} d_{i+1} &= d_i(1 - p_i^*) + Bd_i^{1/2}\ln^{3/2} n \\ &= d_i\left(1 - e^{-\binom{k}{\ell}}\right)\left(1 + O\big(D_i^{-1/2}\ln^{3/2} n\big) + O\big(d_i^{-1/2}\ln^{3/2} n\big)\right) \\ &= d_i\left(1 - e^{-\binom{k}{\ell}}\right)\left(1 + O\big(D_i^{-1/2}\ln^{2} n\big)\right). \end{aligned} \tag{4.38}$$

Similarly, we have

$$\begin{aligned} D_{i+1} &= D_i(1 - p_i^*)^{\binom{k}{\ell}-1} + BD_i^{1/2}\ln^{3/2} n \\ &= D_i\left(1 - e^{-\binom{k}{\ell}}\right)^{\binom{k}{\ell}-1}\left(1 + O\big(D_i^{-1/2}\ln^{3/2} n\big)\right). \end{aligned} \tag{4.39}$$

Set $\lambda = \left(1 - e^{-\binom{k}{\ell}}\right)$, and $t = \binom{k}{\ell} - 1$. It follows from (4.38) and (4.39), that

$$d_i = d_0\,\lambda^i\left(1 + O\big(D_0^{-1/2}\ln^2 n\big)\right)\cdots\left(1 + O\big(D_{i-1}^{-1/2}\ln^2 n\big)\right), \tag{4.40}$$

$$D_i = D_0\,\lambda^{it}\,\underbrace{\left(1 + O\big(D_0^{-1/2}\ln^2 n\big)\right)\cdots\left(1 + O\big(D_{i-1}^{-1/2}\ln^2 n\big)\right)}_{E}. \tag{4.41}$$

To estimate the error term E in both (4.40) and (4.41) above, let $c > 0$ be a constant such that

$$E \geq \left(1 - cD_0^{-1/2}\ln^2 n\right)\ldots\left(1 - cD_{i-1}^{-1/2}\ln^2 n\right) \tag{4.42}$$

and

$$E \le \left(1 + cD_0^{-1/2}\ln^2 n\right)\dots\left(1 + cD_{i-1}^{-1/2}\ln^2 n\right). \tag{4.43}$$

Even though the number of big-O terms being multiplied in E is not constant, they all come from (4.38) (or (4.39)), hence the choice of c is licit. Therefore, using (4.43) and the exponential bound $1 + x \le e^x$, we obtain

$$E \le \exp\left\{c\ln^2 n \sum_{j=0}^{i-1} D_j^{-1/2}\right\}. \tag{4.44}$$

To bound the sum of $D_j^{-1/2}$, we observe that (4.39) implies

$$D_i = D_{i-1}\lambda^t(1 + o(1)) \le D_{i-1}\lambda^{t-1},$$

which can be generalized, for every $0 \le j < i \le s$, as

$$D_i \le D_j\lambda^{(i-j)(t-1)}. \tag{4.45}$$

Using (4.45), and the assumption that $D_i > \ln^5 n$, for $i < s$, we have

$$\begin{aligned}\sum_{j=0}^{i-1} D_j^{-1/2} &\le D_{i-1}^{-1/2} + D_{i-2}^{-1/2} + \cdots + D_0^{-1/2}\\ &\le D_{i-1}^{-1/2}\left(1 + \left(\lambda^{\frac{1}{2}(t-1)}\right) + \cdots + \left(\lambda^{\frac{1}{2}(t-1)}\right)^{(i-1)}\right)\\ &\le D_{i-1}^{-1/2}\left(1 - \lambda^{\frac{1}{2}(t-1)}\right)^{-1}.\end{aligned} \tag{4.46}$$

Now we use (4.46) to bound E from above as follows.

$$\begin{aligned}E &\le \exp\left\{c\ln^2 n\, D_{i-1}^{-1/2}\left(1 - \lambda^{\frac{1}{2}(t-1)}\right)^{-1}\right\}\\ &\le \exp\left\{c\left(1 - \lambda^{\frac{1}{2}(t-1)}\right)^{-1}\ln^{-1/2} n\right\}.\end{aligned}$$

Since c is positive, and $1 + 2x \ge e^x$ for $0 \le x \le 1/2$, we obtain

$$E \le 1 + o(1).$$

Similarly, we obtain a lower bound on E. By (4.42), and $1+2x \geq e^x$ for $0 \leq x \leq 1/2$, it follows that

$$\begin{aligned} E &\geq \exp\Big\{-\frac{c}{2}\ln^2 n \sum_{j=0}^{i-1} D_j^{-1/2}\Big\} \\ &\geq \exp\Big\{-\frac{c}{2}\ln^2 n\, D_{i-1}^{-1/2}\big(1-\lambda^{\frac{1}{2}(t-1)}\big)^{-1}\Big\} \qquad (4.47) \\ &\geq \exp\Big\{-\frac{c}{2}\big(1-\lambda^{\frac{1}{2}(t-1)}\big)^{-1}\ln^{-1/2} n\Big\} \qquad (4.48) \\ &= 1+o(1). \end{aligned}$$

Hence, we can restate (4.40) and (4.41) as

$$d_i = d_0\,\lambda^i(1+o(1)), \qquad (4.49)$$

$$D_i = D_0\,\lambda^{it}(1+o(1)). \qquad (4.50)$$

We now compute s and d_s. Observe that $D_s \leq \ln^5 n < D_{s-1}$ implies

$$\lambda^{st}D_0(1+o(1)) \leq \ln^5 n < \lambda^{(s-1)t}D_0(1+o(1)),$$

which implies

$$\lambda^s = \Theta\Big(\frac{\ln^5 n}{D_0}\Big)^{1/t}. \qquad (4.51)$$

Making $e^{-bs} = \lambda^s$, we get

$$-bs = \Theta(\ln\ln n) - \Theta(\ln D_0).$$

The assumption that $D_0 \leq 2n\ln n$ yields

$$s = \Theta(\ln n). \qquad (4.52)$$

Using (4.51) and (4.49), together with the assumption that $d_0 = n \leq D_0$, we obtain

$$\begin{aligned} d_s &= \lambda^s d_0(1+o(1)) \\ &= O\Big(\frac{\ln^5 n}{D_0}\Big)^{1/t} d_0 \\ &= O\Big(n^{1-1/t}\ln^{4/t} n\Big). \qquad (4.53) \end{aligned}$$

Now we bound the number of ℓ-shadows containing a fixed $(\ell-1)$-subset $M \in \binom{V}{\ell-1}$ that were wasted along the process, after s bites. At each bite i, the number of such ℓ-shadows is bounded by $21Ad_iD_i^{-1/2}\ln^{3/2} n$. Therefore, the number of ℓ-shadows containing M that were wasted along the s iterations can be bounded, using (4.49), (4.50), and the assumption that $n \le D_0$, as follows.

$$\begin{aligned}\sum_{i=0}^{s-1} 21Ad_iD_i^{-1/2}\ln^{3/2} n &= 21A\ln^{3/2} n \sum_{i=0}^{s-1} d_iD_i^{-1/2} \\ &= 21A\ln^{3/2} n \sum_{i=0}^{s-1} \big(d_0\lambda^i(1+o(1))\big)\big(D_0\lambda^{it}(1+o(1))\big)^{-1/2} \\ &= O\big(n^{1/2}\ln^{3/2} n\big)\sum_{i=0}^{s-1}\lambda^{i(1-t/2)} \\ &= O\big(n^{1/2}\ln^{3/2} n\,\lambda^{s(1-t/2)}\big).\end{aligned}$$

Using the estimate on λ^s given by (4.51), we conclude that the number of wasted ℓ-shadows containing M is

$$O\left(n^{1/2}\ln^{3/2} n\left(\frac{\ln^5 n}{D_0}\right)^{1/t-1/2}\right) = O\big(n^{1-1/t}\ln^{\gamma} n\big). \tag{4.54}$$

4.8 Proof of Theorem 4.1

We are now ready to prove our main theorem. We are going to show that the desired object, i.e. a locally nearly perfect (n,k,ℓ)-system, can be obtained by taking the union of the selected k-blocks along the s good bites from the previous section. Let $\mathcal{F}$ be the family of k-sets defined by

$$\mathcal{F} = \mathcal{S}_1 \cup \mathcal{S}_2 \cup \cdots \cup \mathcal{S}_s.$$

Clearly, two k-blocks in the same family $\mathcal{S}_i$ do not share ℓ elements. Recall that $\mathcal{S}_i \subseteq \mathcal{H}_{i-1}$. Moreover, since $\mathcal{H}_i \cap \nabla_k(\Delta_\ell(\mathcal{S}_j)) = \emptyset$ for all $j \le i$, no two members of $\mathcal{F}$ intersect in more than $\ell - 1$ elements. Below, we verify that $\mathcal{F}$ is indeed the desired (n, k, ℓ)-system.

Let M be an arbitrary $(\ell-1)$-element subset of $[n]$. Let $\mathcal{H}_i$ be as defined in the previous section. We know

$$\mathcal{H}_0 \supseteq \mathcal{H}_1 \supseteq \mathcal{H}_2 \supseteq \cdots \supseteq \mathcal{H}_s. \tag{4.55}$$

Note that, if an ℓ-shadow L belongs to $\Delta_\ell(\mathcal{H}_i) \setminus \Delta_\ell(\mathcal{H}_{i+1})$, then either $L \in \Delta_\ell(\mathcal{S}_i)$ or $L \in \mathscr{W}_i$ or both. Since every ℓ-subset of $[n]$ is an ℓ-shadow of $\mathcal{H}_0$, the number of uncovered ℓ-subsets containing M is at most

$$|\nabla_\ell(M) \cap \Delta_\ell(\mathcal{H}_s)| + \Big|\bigcup_{i=1}^{s} \nabla_\ell(M) \cap \mathscr{W}_i\Big|. \tag{4.56}$$

Th first term is at most d_s, which by (4.53) is $O\big(n^{1-1/t} \ln^{4/t} n\big)$. The second is $O\big(n^{1-1/t} \ln^{5/t} n\big)$ by (4.54). Hence, the number of uncovered ℓ-subsets containing M is of order $O\big(n^{1-\beta} \ln^{\gamma} n\big)$, for some γ which depends only on k and ℓ.

Appendix

Proposition 5.1 *Let $0 \leq \ell \leq k \leq n$. Then the following binomial identity holds.*

$$\binom{n}{k}\binom{n-\ell}{k-\ell}^{-1} = \binom{n}{\ell}\binom{k}{\ell}^{-1}. \tag{5.1}$$

Proof. We expand the left-hand-side of (5.1) as

$$\Big(\frac{n}{k}\frac{n-1}{k-1}\cdots\frac{n-k+1}{1}\Big)\Big(\frac{k-\ell}{n-\ell}\frac{k-\ell-1}{n-\ell-1}\cdots\frac{1}{n-k+1}\Big).$$

Canceling terms yields

$$\frac{n}{k}\frac{n-1}{k-1}\cdots\frac{n-\ell+1}{k-\ell+1}.$$

Multiply and divide by $\ell!$, then regroup terms to get the right-hand-side of (5.1). □

Proposition 5.2 *For a positive integer ℓ and a real number $x \leq 1$, the following inequality holds.*

$$(1-x)^{\ell} \geq (1-\ell x). \tag{5.2}$$

Proof. The proof is by induction on ℓ. If $\ell = 1$ the inequality holds trivially. Suppose that $\ell > 1$, and assume inequality 5.2 holds for smaller values of ℓ. Using the induction hypothesis, we obtain

$$\begin{aligned}
(1-x)^{\ell} &= (1-x)(1-x)^{\ell-1} \\
&\geq (1-x)(1-(\ell-1)x) \\
&= 1-\ell x+(\ell-1)x^2 \\
&\geq 1-\ell x.
\end{aligned}$$

The proposition follows. $\square$

Proposition 5.3 *For a positive integer D the following inequality holds.*

$$\left(1-\frac{1}{D}\right)^{D-1} \geq \mathrm{e}^{-1}.$$

Proof. Since

$$\lim_{x\to\infty}\left(1-\frac{1}{x}\right)^{x} = \lim_{x\to\infty}\left(1-\frac{1}{x}\right)^{x-1} = \mathrm{e}^{-1},$$

it is enough to show that for any integer $D > 1$, we have

$$\left(1-\frac{1}{D+1}\right)^{D} < \left(1-\frac{1}{D}\right)^{D-1}. \tag{5.3}$$

Observe that (5.3) is equivalent to

$$\left(1-\frac{1}{D+1}\right) < \left(\frac{1-\frac{1}{D}}{1-\frac{1}{D+1}}\right)^{D-1}. \tag{5.4}$$

In fact, this inequality is true since the right-hand-side of (5.4) can be bounded as follows.

$$\begin{aligned}
\left(\frac{1-\frac{1}{D}}{1-\frac{1}{D+1}}\right)^{D-1} &= \left(1-\frac{1}{D^2}\right)^{D-1} \\
&> \left(1-\frac{(D-1)}{D^2}\right) \\
&> \left(1-\frac{(D-1)}{(D-1)(D+1)}\right) \\
&= \left(1-\frac{1}{D+1}\right),
\end{aligned}$$

where the the first inequality follows from Proposition 5.2. This completes the proof. $\square$

Bibliography

[1] J. Albree and S. H. Brown. “a valuable monument of mathematical genius”: The ladies’ diary (17041840). *Historia Math.*, 36(1):10–47, 2009.

[2] Noga Alon, Jeong-Han Kim, and Joel Spencer. Nearly perfect matchings in regular simple hypergraphs. *Israel J. Math.*, 100:171–187, 1997.

[3] Noga Alon and Joel H. Spencer. *The probabilistic method*. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience [John Wiley & Sons], New York, second edition, 2000. With an appendix on the life and work of Paul Erdős.

[4] Noga Alon and Raphael Yuster. On a hypergraph matching problem. *Graphs Combin.*, 21(4):377–384, 2005.

[5] R. H. Bruck and H. J. Ryser. The nonexistence of certain finite projective planes. *Canadian J. Math.*, 1:88–93, 1949.

[6] Peter J. Cameron. *Combinatorics: topics, techniques, algorithms*. Cambridge University Press, Cambridge, 1994.

[7] Charles J. Colbourn and Jeffrey H. Dinitz, editors. *The CRC handbook of combinatorial designs*. CRC Press Series on Discrete Mathematics and its Applications. CRC Press, Boca Raton, FL, 1996.

[8] P. Erdős and H. Hanani. On a limit theorem in combinatorial analysis. *Publ. Math. Debrecen*, 10:10–13, 1963.

[9] P. Frankl and V. Rödl. Near perfect coverings in graphs and hypergraphs. *European J. Combin.*, 6(4):317–326, 1985.

[10] Haim Hanani. On quadruple systems. *Canad. J. Math.*, 12:145–157, 1960.

[11] Haim Hanani. The existence and construction of balanced incomplete block designs. *Ann. Math. Statist.*, 32:361–386, 1961.

[12] D. R. Heath-Brown and H. Iwaniec. On the difference between consecutive primes. *Invent. Math.*, 55(1):49–69, 1979.

[13] Svante Janson, Tomasz Łuczak, and Andrzej Rucinski. *Random graphs*. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2000.

[14] T. P. Kirkman. On a problem in combinations. *Cambridge and Dublin Math. J.*, 2:191–204, 1847.

[15] A. V. Kostochka, V. Rödl, and L. A. Talysheva. On systems of small sets with no large Δ-subsystems. *Combin. Probab. Comput.*, 8(3):265–268, 1999.

[16] Nikolai N. Kuzjurin. On the difference between asymptotically good packings and coverings. *European J. Combin.*, 16(1):35–40, 1995.

[17] László Lovász. *Combinatorial problems and exercises*. AMS Chelsea Publishing, Providence, RI, second edition, 2007.

[18] Nicholas Pippenger and Joel Spencer. Asymptotic behavior of the chromatic index for hypergraphs. *J. Combin. Theory Ser. A*, 51(1):24–42, 1989.

[19] M. Reiss. Ueber eine steinersche kombinatorische aufgabe. *Crelle's Journal*, 65:326, 1859.

[20] Vojtěch Rödl. On a packing and covering problem. *European J. Combin.*, 6(1):69–78, 1985.

[21] Vojtěch Rödl and Jan Zich. Triangulations and the Hajós conjecture. *J. Graph Theory*, 59(4):293–325, 2008.

[22] Van H. Vu. New bounds on nearly perfect matchings in hypergraphs: higher codegrees do help. *Random Structures Algorithms*, 17(1):29–63, 2000.

[23] Richard M. Wilson. An existence theory for pairwise balanced designs. I. Composition theorems and morphisms. *J. Combinatorial Theory Ser. A*, 13:220–245, 1972.

[24] Richard M. Wilson. An existence theory for pairwise balanced designs. II. The structure of PBD-closed sets and the existence conjectures. *J. Combinatorial Theory Ser. A*, 13:246–273, 1972.

[25] Richard M. Wilson. An existence theory for pairwise balanced designs. III. Proof of the existence conjectures. *J. Combinatorial Theory Ser. A*, 18:71–79, 1975.

[26] V. A. Zinoviev. Cascade equal-weight codes and maximal packings. *Problems Control Inform. Theory/Problemy Upravlen. Teor. Inform.*, 12(1):3–10, 1983.