Economical covers with geometric applications

Noga Alon [*] Béla Bollobás [†] Jeong Han Kim [‡] Van H. Vu [§]

Abstract

Given a hypergraph H, a *cover* of H is a collection of edges whose union is the set of vertices; the minimal number of edges in a cover is the *covering number* $\mathrm{cov(H)}$ of H. The *maximal codegree* $\Delta_2(H)$ is the maximal number of edges containing two fixed vertices of H. For $D = 1, 2, \ldots$, let H_D be a D-regular k-uniform hypergraph on n vertices, where k and n are functions of D. Among other results, we shall prove that if $\Delta_2(H_D) = o(D/e^{2k} \log D)$ and $k = o(\log D)$ then $\mathrm{cov(H_D) = (1 + o(1))n/k}$; this extends the known result that this holds for fixed k. On the other hand, if $k \geq 4 \log D$ then $\mathrm{cov(H_D)} \geq \Omega(\frac{\mathrm{n}}{\mathrm{k}} \log(\frac{\mathrm{k}}{\mathrm{\log D}}))$ may hold even when $\Delta_2(H_D) = 1$. Several extensions and variants are also obtained, as well as the following geometric application. The minimum number of lines required to separate n random points in the unit square is, almost surely, $\Theta(n^{2/3}/(\log n)^{1/3})$.

1 Introduction

A *hypergraph* $H = (V, E)$ consists of a *vertex set* V, and a collection E of subsets of V, called *edges*. We say that a hypergraph is *k-uniform* if all edges have cardinality k. Given a hypergraph H, the *degree* $d_H(v)$ of a vertex v is the number of edges containing v, and H is *D-regular* if the degree of every vertex is exactly D. Given two vertices u and v, the *codegree* $d_H(u, v)$ of u and v is the number of edges containing both u and v; we write $\Delta_2(H)$ for the *maximal codegree* of H. In the rest of this section, we consider only D-regular k-uniform hypergraphs on n vertices, and all statements are made only for this class of hypergraphs, unless otherwise specified. We assume that D tends to infinity, and the asymptotic notation such as o, O, Ω, is used under this assumption. All logarithms are to base e.

A *cover* of a hypergraph H is a set of edges whose union is the vertex set of H; thr *covering number* $\mathrm{cov(H)}$ of H is the minimal number of edges in a cover. It is clear that if H is k-uniform and has n vertices, then $\mathrm{cov(H) \geq n/k}$. A cover $\mathbf{C}$ is *nearly optimal* if $|\mathbf{C}| = \frac{n}{k}(1 + o(1))$. A closely related notion

[*]Department of Mathematics, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv, Israel. Research supported in part by a USA-Israeli BSF grant, by the Israel Science Foundation and by the Hermann Minkowski Minerva Center for Geometry at Tel Aviv University. Email: noga@math.tau.ac.il. Part of this research was performed during a visit at Microsoft Research.

[†]Department of Mathematical Sciences, University of Memphis, Memphis, TN 38152 and Trinity College, Cambridge CB2 1TQ, England. Email: bollobas@msci.memphis.edu and b.bollobas@dpmms.cam.ac.uk Research supported in part by NSF grant no. DMS-9971788.

[‡]Microsoft Research, Redmond, WA 98052. Email: jehkim@microsoft.com

[§]Microsoft Research, Redmond, WA 98052. Email: vanhavu@microsoft.com

is that of a *nearly perfect matching.* A matching in H is a collection of pairwise disjoint edges. Given a matching $\mathbf{M}$ in H, we denote by $U(\mathbf{M})$ the number of vertices left uncovered by $\mathbf{M}$, and we say that $\mathbf{M}$ is nearly perfect if $U(\mathbf{M}) = o(n)$. It is obvious that if k is fixed and H has a nearly perfect matching, then it also has a nearly optimal cover. The question whether a hypergraph has a nearly perfect matching or a nearly optimal cover (and how close to optimal they are) is among the central questions in extremal combinatorics, and has been investigated intensively. Pippenger [23], based on an original work of Rödl [25] on the Erdős-Hanani conjecture, showed that the existence of a nearly perfect matching can be guaranteed by a simple condition on the codegrees. He proved that if k is fixed and $\Delta_2(H) = o(D)$ then H contains a nearly perfect matching. (Recall that we always assume that the hypergraph is k-uniform and D-regular). In recent years, there have been several attempts to make Pippenger's observation quantitatively sharper, namely, to give a better bound on the number of uncovered vertices. Alon, Kim and Spencer [3] showed that if the hypergraph is *simple*, that is, the codegrees are at most 1, then there is a matching $\mathbf{M}$ such that $U(\mathbf{M}) = O(\frac{n}{D^{1/(k-1)}})$, with an additional polylog$D$ factor for $k = 3$. Grable [10] and Kostochka and Rödl [21] proved a related result for hypergraphs with moderate codegrees. The most recent result in this direction, proved by Vu [27], implies that, for any constant $k \geq 4$, if $\Delta_2(H) \leq C$ then the hypergraph H contains a matching M such that $U(M) = O(n(\frac{C}{D})^{1/(k-1)} polylog D)$. This result can also be stated in terms of covering as follows.

Theorem 1.1 *Let $k \geq 4$ be a fixed positive integer. If H is a D-regular k-uniform hypergraph on n vertices with maximum codegree C, then*

$$\mathrm{cov(H)} = \frac{\mathrm{n}}{\mathrm{k}} + \mathrm{O}(\mathrm{n}(\frac{\mathrm{C}}{\mathrm{D}})^{1/(\mathrm{k}-1)} \log^{\gamma} \mathrm{D}),$$

where $\gamma = 3/2$ if $k = 4$ and $\gamma = 1$ if $k > 4$.

As already mentioned, the results above were proved with k constant and $D \to \infty$. ¿From these results it is not clear what happens if k also tends to infinity as a function of D, say, $k = \log^{1/2} D$ or D^{ϵ}.

The first goal of this paper is to study the behaviour of the optimal cover of a hypergraph in the case of non-constant k. A simpler version of the main theorems in the first part of the paper is the following result, which extends Theorem 1.1 to non-constant k and also gives a more precise logarithmic term. The full version is presented in Section 3 (see Theorem 3.7).

Theorem 1.2 *Let $\mathcal{H}$ be a D-regular k-uniform hypergraph with maximum codegree C. Then*

$$\mathrm{cov(H)} = \frac{\mathrm{n}}{4}\Big(1 + \mathrm{O}\Big(\Big(\frac{(\log \mathrm{D})^4\mathrm{C}}{\mathrm{D}}\Big)^{1/3}\Big)\Big)$$

if $k = 4$, and

$$\mathrm{cov(H)} = \frac{\mathrm{n}}{\mathrm{k}}\Big(1 + \mathrm{O}\Big(\Big(\frac{\mathrm{C}\log(1+\mathrm{C})}{\mathrm{D}}\Big)^{1/(\mathrm{k}-1)}\Big)\Big)$$

if $k > 4$ and $e^{2k}C = o(D/\log D)$.

If $4 < k = o\Big(\log \frac{D}{C\log(1+C)}\Big)$, then the error term $O\Big(\Big(\frac{C\log(1+C)}{D}\Big)^{1/(k-1)}\Big)$ is $o(1)$ and in this case H has a nearly optimal cover. Thus Theorem 1.2 has the following consequence.

Corollary 1.3 *Let $\mathcal{H}$ be a D-regular k-uniform hypergraph with maximum codegree C. If $4 < k = o\Big(\log \frac{D}{C\log(1+C)}\Big)$ then $\mathcal{H}$ has a cover of size $\frac{(1+o(1))n}{k}$.*

In case k does not satisfy the condition $e^{2k}C = o(D/\log D)$, one can still derive the following bound from Theorem 1.2 by simply splitting every edge into smaller edges (for more details, see the argument following Theorem 3.7).

Corollary 1.4 *Let $\mathcal{H}$ be a D-regular k-uniform hypergraph with maximum codegree C. If $C = o\Big(\frac{D}{\log D}\Big)$, and $k \geq (1/3)\log\frac{D}{C\log D}$, then $\mathcal{H}$ has a cover of size*

$$O\Big(\frac{|V|}{\log(D/(C\log D))}\Big).$$

Theorem 1.2 also yields the following result on nearly perfect matchings.

Corollary 1.5 *Let $\mathcal{H}$ be a D-regular k-uniform hypergraph. Then there is a matching which covers all but*

$$O\Big(\Big(\frac{(\log D)^4 C}{D}\Big)^{1/3} n\Big)$$

vertices when $k = 4$, and all but

$$O\Big(k\Big(\frac{C\log(1+C)}{D}\Big)^{1/(k-1)} n\Big)$$

if $k > 4$ and $e^{2k}C = o(D/\log D)$.

In particular, if $k > 4$ and

$$k = o\Big(\log\frac{D}{C\log(1+C)}\Big),$$

then there is a matching of size $\frac{(1-o(1))n}{k}$.

Theorem 1.2 implies, for instance, that for any simple D-regular k-uniform hypergraph H with $k = o(\log D)$, H always contains a nearly optimal cover. When $k = \Theta(\log D)$, the bound is $O(n/k)$. In Section 2, we show that for $k = \lfloor c\log D \rfloor$ with $c > 4$, one can construct a simple k-uniform hypergraph whose best cover has size at least $\Omega(\frac{n}{k}\log c)$. Therefore, in the worst case, the ratio between the cardinality of the smallest cover and n/k tends to infinity as c tends to infinity. This, in a sense, means that $k = \log D$ is the threshold of the property of containing a nearly optimal cover.

In the second part of the paper we describe two geometric consequences of the above theorem. The first one is a determination, up to a constant factor, of the typical minimum number of lines needed to cover a random set of points of the projective plane $PG(2,q)$ obtained by picking each point with probability p. This is done for all admissible values of p, and turns out to be a rather simple consequence of the theorm above. For the special case $p \geq b\log q/q$ for a sufficiently large constant b, a more precise estimate was given by Kahn in [13].

The second application is more complicated and deals with the typical minimum number of lines needed to separate a set of n random points in the unit square. A collection of lines separates a set of points if no two points in the set lie in the same connected component of the complement of the collection. We show that the minimum size of a collection of lines that separates a set of n random points in the unit square is, almost surely, $\Theta(n^{2/3}/(\log n)^{1/3})$.

The rest of the paper is organized as follows. In Section 2 we show that if k is asymptotically bigger than $\log D$ then there are simple D-regular k-uniform hypergraphs with no nearly optimal covers. Section 3 contains the proof of Theorem 1.2 and some extensions. The two geometric applications are described in Sections 4 and 5, and the final Section 6 contains some concluding remarks and open problems.

2 Simple regular uniform hypergraphs with no small covers

In this section we prove that if $D \geq k \geq 3$ and $k \geq 4 \log D$, then there are simple D-regular k-uniform hypergraphs on n vertices in which any set of edges that covers all vertices contains at least $\Omega(\frac{n}{k} \log(\frac{k}{\log D}))$ edges. This fact, as well as some related results, is proved by a probabilistic construction. We describe a probabilistic procedure for generating D-regular k-uniform hypergraphs on n vertices. This is the hypergraph analog of the known model for random regular graphs (see [6]). The probability that the hypergraph generated in this procedure is simple is small, but positive. Yet, we show that this probability is much larger than the probability it contains a small cover, implying the desired result. The treatment here can be done in a way similar to the known one for graphs, but we prefer to describe a simple, self-contained proof whose advantage is that it can be used to obtain some conclusions even when n is not bigger than a polynomial in k and D.

The actual details require some notation and lemmas, as follows. For two integers $n \geq D \geq 3$, let $V = V(n, D)$ denote the set of all ordered pairs (i, j) with $0 \leq i < n$ and $1 \leq j \leq D$. For an integer $k \geq 3$, a permutation $\sigma = \sigma_1, \sigma_2, \ldots, \sigma_{nD}$ of V is called k-legal if for every $0 \leq i < j < n$ there is at most one pair of elements σ_t and σ_g of σ the distance between which satisfies $|t - g| < k$ and whose first coordinates are i and j respectively.
Note that if a permutation σ of V as above is k-legal, then there are no two members of σ with the same first coordinate the distance between which in σ is at most $2k - 2$ (since in this case there will always be another member of σ whose distance from each of them is less than k, violating the condition that σ is k-legal). Note also that if nD is a multiple of k then any k-legal permutation σ as above, can be used to generate a simple D-regular k-uniform hypergraph $H = H(\sigma)$ on n vertices as follows. The n vertices are $\{0, 1, \ldots, n-1\}$ and the edges are obtained by splitting the permutation into nD/k blocks of k consecutive elements each, and by defining each set of the k first coordinates of the elements in each block to be an edge. As the permutation is k-legal, each edge indeed contains k distinct vertices, and no two edges have more than one vertex in common, as needed. Our model for generating randomly a D-regular k-uniform hypergraph is thus as follows; let σ be a random permutation of V, and let $H = H(\sigma)$ be the corresponding hypergraph. Our objective is to show that with positive (though small) probability this hypergraph is simple and contains no small covers.

We first show that if n is sufficiently large as a function of k and D, then there are some k-legal permutations of $V(n,D)$. This is done in the following lemma. We note that it is not difficult to improve the estimate it provides (and drop the assumption that n is a prime), but since this requires a somewhat more complicated proof and the present estimate suffices for our purpose, we present only the short proof below.

Lemma 2.1 *For every prime n satisfying $n \geq k^4 D^2$ there exists a k-legal permutation of $V(n,D)$.*

Proof: Let $a_1, a_2, \ldots, a_D$ be D residues modulo n chosen randomly, uniformly and independently. We claim that with positive probability for every $1 \leq i \leq j \leq D$,

$$s_1 a_i + s_2 a_{i+1} \not\equiv \pm(r_1 a_j + r_2 a_{j+1}) (\ mod\ n), \tag{1}$$

for all nonnegative integers s_1, s_2, r_1, r_2 satisfying $1 \leq s_1 + s_2 \leq k-1$ and $1 \leq r_1 + r_2 \leq k-1$, unless the equality holds trivially (that is $i = j$ and $(s_1, s_2) = (r_1, r_2)$ or $i + 1 = j, s_1 = r_2 = 0$ and $s_2 = r_1$).

To prove this claim note that there are $\binom{D+1}{2} < D^2$ possibilities to choose i and j, and $2[\binom{k+1}{2} - 1]^2 \leq k^4$ possibilities to select s_1, s_2, r_1, r_2 and a sign. For each such fixed selection, the probability that (1) holds is $1/n$ and the claim thus follows by the assumption that $n \geq k^4 D^2$. Given the residues $a_1, a_2 \ldots, a_D$ satisfying the assertion of the claim, define the following permutation of $v(n,D)$:

$$(a_1, 1), (2a_1, 1), \ldots, (na_1, 1)\ (\ = (0,1)\ ), (a_2, 2), (2a_2, 2), \ldots, (na_2, 2)\ (\ = (0,2)\ ),$$

$$\ldots, (a_D, D), (2a_D, D), \ldots, (na_D, D).$$

Note that the difference, modulo n, between the first coordinates of any two elements whose distance in the permutation is at most $(k-1)$ is of the form $s_1 a_i + s_2 a_{i+1}$, where $s_1, s_2 \geq 0$ and $1 \leq s_1 + s_2 \leq k-1$, and hence the permutation is k-legal, by (1). □

Lemma 2.2 *Let $f(n,k,D)$ denote the number of k-legal permutations on $v(n,D)$. If $n \geq 4(k-1)^2 D$ then*

$$f(n+1,k,D) \geq$$

$$f(n,k,D)(nD+1)(nD+2-4(k-1)^2 D)(nD+3-2\cdot 4(k-1)^2 D) \cdots (nD+D-(D-1)4(k-1)^2 D).$$

Proof: Each k-legal permutation on $v(n,D)$ can be completed in several ways to a k-legal permutation on $v(n+1,D)$ by inserting the elements $(n,1), (n,2), \ldots, (n,D)$ one by one, keeping the permutation k-legal, as follows. There are $nD+1$ possible places to insert $(n,1)$. After $(n,1), \ldots, (n,i)$ have already been placed, there are at most $4(k-1)^2 iD$ forbidden places to insert $(n, i+1)$, as it cannot be within distance $k-1$ from any element whose first coordinate is the first coordinate of some other element that lies within distance $k-1$ of some of the i elements (n,j) that have already been inserted. This supplies the desired estimate. □

Since for every s there is a prime between s and $2s$, the two last lemmas imply the following.

Corollary 2.3 *For $n \geq 2k^4D^2$, the probability that a random permutation of $v(n, D)$ is k-legal is at least*

$$\frac{1}{(2k^4D^2)^{2k^4D^2}} e^{-4k^2D^2 \log n}.$$

□

Next we estimate the probability that the random hypergraph constructed by our permutation contains a small cover. For simplicity, we omit all floor and ceiling signs whenever these are not crucial. We first prove a simple correlation inequality, a special case of which is applied later. Define $N = \{0, 1, 2, \dots, n-1\}$ and let D_i, $0 \leq i < n$ be positive integers. Let $U = U(n, D_1, \dots, D_{n-1})$ denote the set of all $\sum_{i \in N} D_i$ elements (i, b_i) where $i \in N, 1 \leq b_i \leq D_i$, and let τ be a random permutation of U. (Note that in the special case $D_i = D$ for all i, the set U is simply $v(n, D)$ and hence in this case τ is the random permutation used in the definition of our hypergraph.) Let $t \leq \sum_{i \in N} D_i$ be a positive integer. We say that i *appears in the t-prefix of τ* if there is an element among the first t members of τ whose first coordinate is i. Let A_i denote this event, and let $P_N(A_i)$ denote its probability. Similarly, for $S \subset N$ and $j \notin S$, let $P_N(\cap_{i \in S} A_i)$ denote the probability of the event that all members of S appear in the t-prefix of τ, and let $P_{N-j}(\cap_{i \in S} A_i)$ denote the probability of the event that all members of S appear in the t prefix of the permutation obtained from τ by deleting all elements whose first coordinate is j. Obviously, the latter event contains the former one, implying the following:
Fact:

$$P_N(\cap_{i \in S} A_i) \leq P_{N-j}(\cap_{i \in S} A_i).$$

Lemma 2.4 *In the above notation, for every $S \subset N$,*

$$P_N(\cap_{i \in S} A_i) \leq \prod_{i \in S} P_N(A_i).$$

Proof: We apply induction on $|S|$. For $|S| = 1$ the result is trivial. Assuming it holds for all sets of size $|S| - 1$, suppose $S \subset N$, $|S| \geq 2$. Choose, arbitrarily, $j \in S$ and put $S' = S - j$. Then

$$P_N(\cap_{i \in S} A_i) = P_N(\cap_{i \in S'} A_i) - P_N(\cap_{i \in S'} A_i \cap \overline{A_j})$$

$$= P_N(\cap_{i \in S'} A_i) - P_N(\overline{A_j}) P_N(\cap_{i \in S'} A_i | \overline{A_j}) = P_N(\cap_{i \in S'} A_i) - P_N(\overline{A_j}) P_{N-j}(\cap_{i \in S'} A_i)$$

$$\leq P_N(\cap_{i \in S'} A_i) - P_N(\overline{A_j}) P_N(\cap_{i \in S'} A_i),$$

where the last inequality follows from the fact above. By the induction hypothesis, the last expression is simply

$$P_N(A_j) P_N(\cap_{i \in S'} A_i) \leq P_N(A_j) \prod_{i \in S'} P_N(A_i) = \prod_{i \in S} P_N(A_i),$$

completing the proof. □

Returning to our random permutation σ of $v(n, D)$, we next prove the following.

Lemma 2.5 *For every fixed $i \in N$, the probability that i does not appear in the nr-prefix of σ is at least $(1 - \frac{nr}{Dn - D + 1})^D$. In particular, for all $D \geq 3$ and, say, $n \geq 100D^2$, this probability is at least 4^{-r}. Moreover, for D and n as above, the probability that all element of N appear in the nr-prefix of σ is at most $e^{-n/4^r}$.*

Proof: The probability that i does not appear in the nr-prefix of σ is precisely

$$\frac{\binom{Dn-nr}{D}}{\binom{Dn}{D}} \geq (\frac{Dn-nr-D+1}{Dn-D+1})^D = (1-\frac{nr}{Dn-D+1})^D \geq 4^{-r},$$

where the last inequality follows from the fact that $n \geq 100D^2$ and $D \geq 3$. Therefore, the probability that i does appear is at most $1-4^{-r}$, and by Lemma 2.4 the probability that all members of N appear is thus at most $(1-4^{-r})^n \leq e^{-n/4^r}$. □

We note that it is possible to give a similar bound for the probability that all but at most, say, $\frac{n}{2\cdot 4^r}$ elements appear in the nr-prefix of σ, but the bound above suffices for our purpose here.

Recall that the permuation σ is partitioned into nD/k blocks of k consecutive elements each, in order to define the edges of the corresponding hypergraph $H = H(\sigma)$. It thus follows, by the last lemma and the obvious fact that there is nothing special about the first nr/k blocks, that the probability that all elements of N appear in any given fixed set of nr/k blocks of σ does not exceed $e^{-n/4^r}$. As there are $\binom{Dn/k}{nr/k}$ possibilities to choose nr/k blocks, it follows that the probability that σ contains some set of $\frac{n}{k}r$ blocks in which all vertices appear does not exceed

$$\binom{Dn/k}{nr/k} e^{-\frac{n}{4^r}} \leq (\frac{eD}{r})^{nr/k} e^{-\frac{n}{4^r}}.$$

Combining this with Corollary 2.3 we obtain the following result.

Theorem 2.6 *Suppose that $n \geq 2k^4D^2$, nD is divisible by k and the following two inequalities hold:*

$$\frac{n}{4^r} > 2\frac{nr}{k}\log(\frac{eD}{r}), \tag{2}$$

$$\frac{n}{4^r} > 2\cdot 4k^2D^2\log n + 2\cdot 2k^4D^2\log(2k^4D^2). \tag{3}$$

Then there is a simple D-regular k-uniform hypergraph with n vertices containing no cover of size $\frac{n}{k}r$.

Proof: Let σ be a random permutation of $v(n, D)$. By the assumptions, Corollary 2.3 and the paragraph preceding the statement of the theorem, with positive probability $H(\sigma)$ is a simple hypergraph that satisfies the required conditions. □

Remarks.

- For every fixed k, D and r, the inequality (3) always holds provided n is sufficiently large, whereas (2) holds provided $k > 2r4^r\log(\frac{eD}{r})$. This implies the following.

  Corollary 2.7 *There exists an absolute positive constant c such that for all $D \geq k \geq 4\log D$ there is a simple D-regular k-uniform hypergraph on n vertices, containing no cover of size smaller than $c\frac{n}{k}\log(\frac{k}{\log D})$.*

  By a special case of the main result in the next section, if $k \leq 4\log D$, then there is always a cover of size at most $O(\frac{n}{k})$.

- The bound in Corollary 2.7 is tight, up to a constant factor, for all k satisfying, say, $k \geq (\log D)^{1.1}$, even without the assumption that the hypergraph is simple, as shown in the following proposition.

  Proposition 2.8 *Every D-regular k-uniform (not necessarily simple) hypergraph $H = (V, E)$ on n vertices contains a cover of size at most $\frac{n(\log k+1)}{k}$.*

  Proof: We follow the argument in [2]. Let F be a random subset of the set of edges of H obtained by picking each member of E, randomly and independently, to lie in F with probability $\log k/D$. Let $Y = Y(F)$ be the set of all vertices not covered by F. Note that by adding to F an arbitrarily chosen edge containing y for each vertex $y \in Y$ we obtain a cover of H of size at most $|F| + |Y|$. Since the expected size of F is $\frac{nD}{k}\frac{\log k}{D} = \frac{n \log k}{k}$, and the expected size of Y is $n(1 - \frac{\log k}{D})^D \leq n/k$, the expected size of the above cover is at most $\frac{n(\log k+1)}{k}$, completing the proof. $\Box$

- By extending the assertion of lemma 2.5 using a martingale inequality we can prove the following extension of Corollary 2.7.

  There exists an absolute positive constant c such that for all $D \geq k \geq 4 \log D$ there is a simple D-regular k-uniform hypergraph H on n vertices, such that every set of at most $c\frac{n}{k}\log(\frac{k}{\log D})$ edges of H does not cover at least $\frac{n}{2 \cdot 4^r}$ vertices.

  Since we do not use this statement in what follows we omit the detailed proof.

- Being slightly more careful in handling the constants in the above proof of Theorem 2.6 it is easy to check that it shows that for large $D \geq k \geq e \log D$ there is a simple D-regular k-uniform hypergraph on n vertices, containing no cover of size smaller than $(1 + o(1))\frac{n}{k}\log(\frac{k}{\log D})$, where the $o(1)$ term tends to 0 as D tends to infinity. In particular, for every $\epsilon > 0$ there is a $\delta > 0$ and D_0 such that if $D > D_0$ and $k > (1+\epsilon)e \log D$, then there is such a hypergraph containing no cover of size less than $(1 + \delta)\frac{n}{k}$.

- The arguments in the derivation of Theorem 2.6 can be modified to deal with matchings instead of covers. The computation in this case is, in fact, easier. Indeed, the probability that no element of N appears more than once in any given set of m blocks of the random permutation σ of $v(n, D)$ is precisely

  $$\frac{Dn(Dn - D)(Dn - 2D) \cdots (Dn - D(mk - 1))}{Dn(Dn - 1)(Dn - 2) \cdots (Dn - mk + 1)}.$$

  There are $\binom{Dn/k}{m}$ possibilities to choose m blocks, and therefore, by immitating the computation in the proof of Theorem 2.6 we obtain the following statement, whose detailed (simple) proof is omitted.

  Proposition 2.9 *There exists an absolute positive constant c such that for all D and k there is a simple D-regular k-uniform hypergraph on n vertices, containing no matching of size bigger than $c\frac{n}{k^2}\log D$.*

3 Economical covers

In this section, we use the *nibble method* to construct the desired cover. The method was first introduced by Ajtai, Komlós and Szemerédi [1] and used by Rödl [25] to confirm a conjecture of Erdős and Hanani [8] regarding Partial Steiner Systems. More sophisticated constructions have been developed during the last decade to solve various intriguing combinatorial problems dealing with hypergraph matching and Partial Steiner System [3, 10, 18, 21, 23, 27], hypergraph edge coloring [24], and list-coloring [14, 15], sparse graph coloring [16, 12, 28, 29], the Ramsey number $R(3,t)$ [17], total coloring [22], and complete arcs in projective planes [19]. The analysis of these constructions is usually based on martingale or isoperimetric inequalities (see e.g. [15, 3, 26, 20]).

The nibble method, which may be regarded as an approximation of the random greedy method, depends centrally on the notion of constructing the desired object in small random increments. To construct a matching of almost optimal size, for example, we first choose a small number of edges so that only few pairs of the chosen edges intersect. Then discard all (both chosen and unchosen) edges that intersect any other chosen edge. The remaining chosen edges are permanently in the matching we are constructing. Applying this procedure repeatedly for the remaining unchosen edges will yield an almost optimal matching if the hypergraph satisfies certain conditions.

As pointed out in [4], constructing a cover is easier than constructing a matching, because a cover does not have to satisfy such a rigid condition as a matching. Thus we do not have to discard any chosen edges, and this makes the analysis easier. To do this, we shall incrementally construct sets of edges $\mathcal{C}$, called *partial covers*, that have certain properties. This sequence of partial covers will be eventually extended to an economical cover. A vertex is said to be *covered* (by $\mathcal{C}$) if it is contained in at least one edge of $\mathcal{C}$. Otherwise, it is *uncovered.* To be extended to an economical cover, $\mathcal{C}$ is required to cover few vertices more than once and the hypergraph induced on the set of uncovered vertices is required to satisfy some properties so that the same method may be applied to extend it to a larger partial cover. We repeat this procedure until only few vertices remain uncovered, and then add for each uncovered vertex an edge containing it.

To explain how partial covers will be constructed, suppose, for simplicity, that the initial hypergraph is D-regular. It turns out that such a strict regularity condition may be relaxed (see (4) below). We construct a random set of edges to which each edge belongs independently with probability θ/D. Then it covers about θ fraction of the vertices. It is also economical if θ is small enough in the sense that only few vertices are covered more than once. The subhypergraph induced on the remaining uncovered vertices has about $n(1-\theta)$ vertices and each vertex has degree about $D(1-\theta)^{k-1}$. Repeat this procedure until only few vertices remain uncovered. Then one edge for each remaining vertex is enough to obtain the desired cover. Making this argument rigorous involves a somewhat technical proof. The main task here is to show that the degrees of the vertices decrease as predicted, with tolerable error terms.

To turn this sketch into a rgorous argument, let $\mathcal{H} = (V, H)$ be a D-regular k-uniform hypergraph satisfying

$$D - f(D) \le d(x) \le D \quad \forall x \in V, \quad \text{and} \quad C := \max_{x,y \in V} cd(x,y) = o(D/\log D) \tag{4}$$

where f is a positive real-valued function of D with $f(D) \leq 0.1D$. We also assume that $k \leq (1/2)\log D$. Choose each edge of $\mathcal{H}$ with probability θ/D for some parameter θ with $k\theta \leq 0.1$, and let $\mathcal{C}$ be the collection of chosen edges. Then the expected size of $\mathcal{C}$ is bounded by $\theta|V|/k$ since there are at most $D|V|/k$ edges. Moreover, the probability of a vertex being covered by $\mathcal{C}$ is at most $1-(1-\theta/D)^D$ and at least $1-(1-\theta/D)^{D-f(D)}$. It turns out that (artificially) making all such probabilities the same is not only convenient but also helps to control the error terms. This technique was introduced in [15] and simplified in [3]. Define

$$p^* = 1 - e^{-\theta(1+10(\frac{C\log D}{\theta D})^{1/2})}$$

and introduce a random set W to which each vertex belongs independently with a certain probability so that

$$\Pr[v \text{ is covered by } \mathcal{C} \text{ or } v \in W] = p^*.$$

(Here the $10(\frac{\log D}{\theta D})^{1/2}$ term is added for the sake of convenience, and W stands for 'waste'.) Putting it slightly differently, $\Pr[v \in W] = p(v)$, where $p(v)$ satisfies

$$(1-\theta/D)^{d(v)}(1-p(v)) = 1 - p^*.$$

This is possible as $1 - p* \leq (1-\theta/D)^{d(v)}$ for all v. It is also easy to check that

$$p(v) \leq 2\theta f(D)/D. \tag{5}$$

We regard vertices in W as covered vertices for a while and add later for each vertex in W an edge containing it.

Lemma 3.1 *Let $\mathcal{H}$ be a hypergraph as described above. If there is a function $g(D) \geq 10k$ such that*

$$f(aD) - af(D) \geq 20(1-a)(g(D)CD\log D)^{1/2} \tag{6}$$

for all $0.9 \leq a \leq 1$, then there is a collection $\mathcal{C}$ of edges and a set W of vertices with the following properties. For $\theta = (g(D))^{-1}$,

$$|\mathcal{C}| \leq \frac{\theta|V|}{k}(1+1/D), \quad |W| \leq \frac{2\theta f(D)|V|}{D}$$

and for the induced sub-hypergraph $\mathcal{H}'$ on the set $V' = V \setminus (W \cup \bigcup_{e\in\mathcal{C}} e)$ of uncovered vertices,

$$|V'| \leq e^{-\theta}|V|, \quad e^{-(k-1)\theta}D - f(e^{-(k-1)\theta}D) \leq d_{\mathcal{H}'}(x) \leq e^{-(k-1)\theta}D$$

for all $x \in V'$.

The proof of Lemma 3.1 is based on the following Azuma-Hoeffding ([5, 11]) type martingale inequality. For more general similar inequalities and their proofs, the reader may consult [3] and [16].

Lemma 3.2 *Let $X_1, ..., X_m$ be independent random variables with* $\Pr[X_i = 0] = 1 - p_i$ *and* $\Pr[X_i = 1] = p_i$. *Let $Y = Y(X_1, ..., X_m)$ be such that*

$$|Y(X_1, ...X_{i-1}, 1, X_{i+1}, ..., X_m) - Y(X_1, ...X_{i-1}, 0, X_{i+1}, ..., X_m)| \le c_i$$

for all $X_1, ...X_{i-1}, X_{i+1}, ..., X_m$, $i = 1, ..., m$. Then for $\sigma^2 = \sum_{i=1}^{n} p_i(1-p_i)c_i^2$ and $\alpha \le 2\sigma^2/\max_i c_i$ we have

$$\Pr[|Y - E[Y]| \ge \alpha] \le 2e^{-\alpha^2/(4\sigma^2)}.$$

The minimum c_i in the lemma is called the *effect* of the variable X_i, or simply the effect of i. In our case, the random variables are indexed by all edges and $p_e = \theta/D$. The effect of an edge e means the effect of the corresponding random variable X_e.

Proof of Lemma 3.1. Let $\mathcal{C}$, W, $\mathcal{H}'$ and p^* be as above with $\theta = (g(D))^{-1}$. It is enough to show that the last inequalities hold with probability at least $e^{-5|V|D^{-4}}$, i.e.

$$\Pr\left[e^{-(k-1)\theta}D - f(e^{-(k-1)\theta}D) \le d_{\mathcal{H}'}(v) \le e^{-(k-1)\theta}D, \ \forall v \in V'\right] \ge e^{-5|V|D^{-4}}, \tag{7}$$

and then each of the the other inequalities holds with probability at least $1 - e^{-5|V|/D^4}/4$.

To prove (7), it is convenient to define the pseudo degree $d^*(v)$ of a vertex v in V (for $v \in V'$ as well as for $v \notin V'$) as the number of edges containing v all of whose vertices except possibly v are in V':

$$d^*(v) = |\{e \in H : v \in e, e \setminus \{v\} \subset V'\}|.$$

Clearly, $d^*(v) = d_{\mathcal{H}'}(v)$ for all $v \in V'$. We show (7) for $d^*(v)$ by estimating the expectation of $d^*(v)$ first, applying Lemma 3.2, and then using the Lovász Local Lemma.

Notice that

$$E[d^*(v)] = \sum_{e \ni v} \Pr[e \setminus \{v\} \subset V'].$$

For each edge e containing v, there are at most $\sum_{w \in e\setminus\{v\}} d_{\mathcal{H}}(w)$ edges containing a vertex in e other than v. On the other hand, as $cd(x, y) \le C$ implies that there are no more than $C\binom{k-1}{2}$ edges containing two vertices in $e \setminus \{v\}$, at least $\sum_{w \in e\setminus\{v\}} d_{\mathcal{H}}(w) - k^2C/2$ edges contain a vertex in e other than v. Since $e \setminus \{v\} \subset V'$ means that all $w \in e \setminus \{v\}$ are uncovered, or equivalently no $w \in e \setminus \{v\}$ is covered by $\mathcal{C}$ nor belongs to W,

$$1 \le \frac{\Pr[e \setminus \{v\} \subset V']}{\prod_{w \in e\setminus\{v\}} (1 - \theta/D)^{d_{\mathcal{H}}(w)} \Pr[w \notin W]} \le (1 - \theta/D)^{-Ck^2/2}.$$

Moreover,

$$(1 - \theta/D)^{d_{\mathcal{H}}(w)} \Pr[w \notin W] = 1 - p^* = e^{-\theta(1 + 10(\frac{C \log D}{\theta D})^{1/2})}$$

and

$$(1 - \theta/D)^{-Ck^2/2} \le 1 + \theta k^2 C/D \le 1 + o\Big(\Big(\frac{k^2\theta C \log D}{D}\Big)^{1/2}\Big)$$

as $\theta k \le 0.1$ and $kC/D = o(1)$. Thus we have

$$1 \le \frac{\Pr[e \setminus \{v\} \subset V']}{e^{-\theta(k-1)(1+10(\frac{C\log D}{\theta D})^{1/2})}} \le 1 + o\Big(\Big(\frac{k^2\theta C \log D}{D}\Big)^{1/2}\Big),$$

and

$$E[d^*(v)] = e^{-\theta(k-1)(1+10(\frac{C\log D}{\theta D})^{1/2})} d_{\mathcal{H}}(v) + o(k(\theta C D \log D)^{1/2}).$$

The effect of an edge e is at most the number of edges containing v that have a non-empty intersection with $e \setminus \{v\}$. Since for each vertex w in $e \setminus \{v\}$ there are at most C edges containing both v and w, the effect of e is at most kC and

$$\sum_e c_e^2 \le kC \sum_e c_e \le kC \sum_e \sum_{e' \ni v} 1(e \cap e' \neq \emptyset) = kC \sum_{e' \ni v} \sum_e 1(e \cap e' \neq \emptyset) \le Ck^2D^2.$$

Let

$$\alpha = 4k(\theta C D \log D)^{1/2} \quad \text{and} \quad \sigma^2 = (1-\theta/D)(\theta/D)Ck^2D^2 \le \theta C k^2 D.$$

As $f(D) \le 0.1D$ and (6) for $a = 0.9$ imply that

$$\frac{C \log D}{\theta D} \le \frac{1}{400}$$

and

$$4k(\theta C D \log D)^{1/2} \cdot kC \le 2(1-\theta/D)\theta C k^2 D,$$

Lemma 3.2 yields

$$\Pr\Big[\, \Big| d^*(v) - E[d^*(v)] \Big| \ge 4k(\theta C D \log D)^{1/2} \Big] \le 2D^{-4}.$$

Let $\mathrm{A}(v)$ be the event

$$\Big| d^*(v) - E[d^*(v)] \Big| \ge 4k(\theta C D \log D)^{1/2}$$

and define a dependency graph making two vertices v, w adjacent if there are three edges e_1, e_2, e_3 such that $v \in e_1, w \in e_3$ and both of $e_1 \cap e_2$, $e_2 \cap e_3$ are not empty. Then $\mathrm{A}(v)$ is mutually independent of all $\mathrm{A}(w)$ with w not adjacent to v. Since v is adjacent to no more than k^2D^3 vertices, the condition of the Lovász Local Lemma, see e.g. [4], $4 \max_v \Pr[\mathrm{A}(v)] k^2 D^3 < 1$ is satisfied and hence

$$\Pr\Big[\, \Big| d^*(v) - E[d^*(v)] \Big| \le 4k(\theta D \log D)^{1/2} \quad \forall v \in V \Big] \ge \prod_v (1 - 2\Pr[\mathrm{A}(v)]) \ge e^{-5|V|D^{-4}}.$$

Since

$$\begin{aligned} E[d^*(v)] &= e^{-\theta(k-1)(1+10(\frac{C\log D}{\theta D})^{1/2})} d_{\mathcal{H}}(v) + o(k(\theta D \log D)^{1/2}) \\ &\le e^{-\theta(k-1)} D - 4k(\theta D \log D)^{1/2} \end{aligned}$$

and

$$\begin{aligned} e^{-\theta(k-1)(1+10(\frac{C\log D}{\theta D})^{1/2})} d_{\mathcal{H}}(v) &\ge e^{-\theta(k-1)(1+10(\frac{C\log D}{\theta D})^{1/2})} (D - f(D)) \\ &\ge e^{-(k-1)\theta} D - 10(k-1)(\theta C D \log D)^{1/2} - e^{-(k-1)\theta} f(D) \\ &\ge e^{-(k-1)\theta} D - f(e^{-(k-1)\theta} D) + 5k(\theta C D \log D)^{1/2}, \end{aligned}$$

the proof of (7) is complete.

We now prove

$$\Pr[|V'| \geq e^{-\theta}|V|] \leq \frac{1}{4}e^{-5|V|/D^4}.$$

The other two inequalities, which are for sums of independent random variables, may be proven similarly and we omit their proofs. The expected size of V' is $(1-p^*)|V| \leq e^{-\theta}|V|(1-1/D)$ by (6). The effect of each edge e is no more than k (for the event $e \in \mathcal{C}$) and the effect of each vertex w is just 1 (for the event $w \in W$), and hence $\sigma^2 \leq k^2(|V|D/k)(\theta/D) + 2\theta|V|f(D)/D \leq 2\theta k|V|$ by (5). Thus

$$\begin{aligned}\Pr[|V'| \geq e^{-\theta}|V|] &\leq \Pr\left[\Big||V'| - E[|V'|]\Big| \geq \frac{|V|}{D}\right] \\ &\leq e^{-|V|/(8\theta k D^2)} \leq \frac{1}{4}e^{-5|V|/D^4}. \qquad \square\end{aligned}$$

To apply Lemma 3.1 recursively, define, for two functions f and g, $\alpha_0 = 1$ and

$$\alpha_i = e^{-\sum_{j=0}^{i-1}\theta_j}, \quad D_i = (\alpha_i)^{k-1}D, \quad \text{and} \quad \theta_i = (g(D_i))^{-1}.$$

If the hypotheses of the lemma are satisfied for all D_i, $i = 0, ..., t$, then we iteratively apply it to have collections $\mathcal{C}_i$ of edges and sets W_i of vertices satisfying

$$|\mathcal{C}_i| \leq \frac{\theta_i|V_i|}{k}(1+1/D_i), \quad |W_i| \leq \frac{2\theta_i f(D_i)|V_i|}{D_i}, \quad \text{and} \quad |V_i| \leq \alpha_i|V|,$$

and for the induced sub-hypergraph $\mathcal{H}_t$ on the set $V_t = V \setminus (\bigcup_{i=0}^{t-1} W_i \cup \bigcup_{i=0}^{t-1}\bigcup_{e\in\mathcal{C}_i} e)$ of uncovered vertices,

$$|V_t| \leq \alpha_t|V|, \quad D_t - f(D_t) \leq d_{\mathcal{H}'}(v) \leq D_t.$$

Adding, for each vertex in $\cup_{i=0}^{t-1}W_i$, an edge containing it, we have a partial cover $\mathcal{D}_t$ that covers all vertices except those in V_t and is of size at most

$$\sum_{i=0}^{t-1}\frac{\theta_i\alpha_i|V|}{k} + \sum_{i=0}^{t-1}\frac{\theta_i(\alpha_i)^{2-k}|V|}{kD} + 2\sum_{i=0}^{t-1}\frac{\theta_i(\alpha_i)^{2-k}f((\alpha_i)^{k-1}D)|V|}{D}.$$

Note that

$$\begin{aligned}\sum_{i=0}^{t-1}\theta_i\alpha_i = \sum_{i=0}^{t-1}\theta_i e^{-\sum_{j=0}^{i-1}\theta_j} &\leq \int_0^{-\log\alpha_t} e^{-x}\,dx + \sum_{i=0}^{t-1}\theta_i(e^{-\sum_{j=0}^{i-1}\theta_j} - e^{-\sum_{j=0}^{i}\theta_j}) \\ &\leq 1 - \alpha_t + 2\sum_{i=0}^{t-1}(\theta_i)^2 e^{-\sum_{j=0}^{i-1}\theta_j}.\end{aligned}$$

Provided

$$g(x) \geq g(y) \quad \text{for } x \geq y \quad \text{and} \quad f((1-\varepsilon)x) \geq f(x)/2, \quad \forall 0 \leq \varepsilon \leq 0.1,$$

one may have

$$\begin{aligned}\sum_{i=0}^{t-1}(\theta_i)^2 e^{-\sum_{j=0}^{i-1}\theta_j} &= \sum_{i=0}^{t-1}\theta_i\Big(g(e^{-(k-1)\sum_{j=0}^{i-1}\theta_j}D)\Big)^{-1}e^{-\sum_{j=0}^{i-1}\theta_j} \\ &\leq 2\int_0^{-\log\alpha_t} e^{-x}\Big(g(e^{-(k-1)x}D)\Big)^{-1}dx \\ &= \frac{2}{(k-1)D^{1/(k-1)}}\int_{D_t}^{D}\frac{x^{1/(k-1)}}{xg(x)}\,dx.\end{aligned}$$

Similarly,

$$\begin{aligned}\frac{1}{D}\sum_{i=0}^{t-1}\theta_i(\alpha_i)^{2-k}\Big(\frac{1}{2k}+f((\alpha_i)^{k-1}D)\Big) &\le \frac{1}{D}\sum_{i=0}^{t-1}\theta_i e^{(k-2)\sum_{j=0}^{i-1}\theta_j}\Big(1+f(e^{-(k-1)\sum_{j=0}^{i-1}\theta_j}D)\Big)\\ &\le \frac{2}{D}\int_0^{-\log\alpha_t} e^{(k-2)x}(1+f(e^{-(k-1)x}D))\,dx\\ &= \frac{2}{(k-1)D^{1/(k-1)}}\int_{D_t}^{D}\frac{x^{1/(k-1)}(1+f(x))}{x^2}\,dx.\end{aligned}$$

Thus we have the following consequence of Lemma 3.1.

Corollary 3.3 *With the same hypotheses as in Lemma 3.1, if*

$$g(x)\ge g(y)\quad \textit{for } x\ge y \quad \textit{and} \quad f((1-\varepsilon)x)\ge f(x)/2,\quad \forall 0\le\varepsilon\le 0.1,$$

then there is a cover $\mathcal{D}_t$ and a subset V_t of V such that $\mathcal{D}_t$ covers all vertices but those in V_t. Furthermore,

$$|\mathcal{D}_t|\le\frac{(1-\alpha_t)|V|}{k}+\frac{2|V|I_{D,k}(f,g:t)}{(k-1)D^{1/(k-1)}},$$

where

$$I_{D,k}(f,g:t)=\int_{D_t}^{D}\frac{x^{1/(k-1)}}{xg(x)}+\frac{x^{1/(k-1)}(1+f(x))}{x^2}\,dx,$$

and for the induced sub-hypergraph $\mathcal{H}_t$ on V_t we have

$$|V_t|\le\alpha_t|V|,\quad D_t-f(D_t)\le d_{\mathcal{H}_t}(v)\le D_t.$$

The partial cover $\mathcal{D}_t$ guaranteed by Corollary 3.3 is nearly optimal and hence contains a large matching. Such a matching may be obtained by discarding all edges containing a vertex covered more than once. Considering the hypergraph $\mathcal{F}$ consisting of all vertices covered by $\mathcal{D}_t$ and all edges in it, the number of discarded edges is at most

$$\sum_{v:d_{\mathcal{F}}(v)\ge 2}d_{\mathcal{F}}(v)\le 2\sum_v(d_{\mathcal{F}}(v)-1)\le 2(k|\mathcal{D}_t|-(1-\alpha_t)|V|)=O\Big(\frac{I_{D,k}(f,g:t)|V|}{D^{1/(k-1)}}\Big).$$

Therefore, Corollary 3.3 has the following immediate consequence.

Corollary 3.4 *With the same hypothesis as in Corollary 3.3, there is a matching that covers all but at most*

$$O\Big(\frac{kI_{D,k}(f,g:t)|V|}{D^{1/(k-1)}}\Big)+\alpha_t|V|$$

vertices.

In the rest of this section, we consider k-uniform hypergraphs $\mathcal{H}$ with maximal codegree C such that for every vertex v we have

$$D-f(D)\le d_{\mathcal{H}}(v)\le D,\quad C=o\Big(\frac{D}{e^{2k}\log D}\Big), \tag{8}$$

where

$$f(x) = 20(x^2 C \log x)^{1/3}, \quad g(x) = \frac{1}{20}\Big(\frac{x}{C\log x}\Big)^{1/3}. \tag{9}$$

In particular, $k \le (1/2)\log D$ since $C \ge 1$. With these functions, we may apply Lemma 3.1 and Corollary 3.3 as long as D_t is larger than a (large) constant times $k^3 C \log(kC)$. Define, for a fixed (large) constant M, $t_0 = t_0(M)$ to be the smallest t such that $D_t \le e^{k-1} MC\log(1+C)$, or equivalently $\alpha_t \le e(MC\log(1+C)/D)^{1/(k-1)}$. It is easy to see that $D_t = o(D)$ by the second condition of (8). It is also routine to check that all conditions are satisfied as long as M is large enough and that $D^{(0)} := D_t \ge 0.9e^{k-1}MC\log(1+C)$. Moreover, for $k > 4$

$$I_{D,k}(f,g;t) = O\Big((C\log(1+C))^{1/(k-1)}\Big).$$

As usual, for each vertex in V_t, we take an edge containing it to find a cover of size

$$\frac{|V|}{4}\Big(1 + O\Big(\Big(\frac{(\log D)^4 C}{D}\Big)^{1/3}\Big)\Big) \quad \text{for } k = 4. \tag{10}$$

If $k > 4$, then since

$$|V_{t_0}| = O\Big(\Big(\frac{C\log(1+C)}{D}\Big)^{1/(k-1)}|V|\Big),$$

there is a cover of size at most

$$\frac{|V|}{k}\Big(1 + O\Big(k\Big(\frac{C\log(1+C)}{D}\Big)^{1/(k-1)}\Big)\Big). \tag{11}$$

Corollary 3.4 also implies that there is a matching that covers all but

$$O\Big(\Big(\frac{(\log D)^4 C}{D}\Big)^{1/3}|V|\Big) \tag{12}$$

vertices if $k = 4$, and all but at most

$$O\Big(k\Big(\frac{C\log(1+C)}{D}\Big)^{1/(k-1)}\Big) \tag{13}$$

vertices if $k > 4$.

If k is large, though (11) is still valid, it may be so weak that it gives only the trivial bound $O(|V|)$. For example, if $k = (1/3)\log D$ and $C = 1$, the bound is just $O(|V|)$. To deal with these cases, we shall show that for the hypergraph $\mathcal{H}^{(0)} := \mathcal{H}_{t_0}$ as in Corollary 3.3, there is a cover of size $O(|V^{(0)}|/k)$ for large k, which yields a cover for $\mathcal{H}$ of size at most

$$\frac{|V|}{k}\Big(1 + O\Big(\Big(\frac{C\log(1+C)}{D}\Big)^{1/(k-1)}\Big)\Big).$$

Lemma 3.5 *For the hypergarph $H^{(0)}$ described above, there is a cover of size $O(|V^{(0)}|/k)$.*

Proof. Take l such that $2^l < k \le 2^{l+1}$ and then randomly discard $k - 2^{l-1}$ vertices from each edge independently of the other edges to create a new edge of size 2^{l-1}. Using the Lovász Local Lemma it is easy to show that with positive probability the degree $d(x)$ of each vertex x of the new hypergraph satisfies

$$\Big|d(x) - 2^{l-1} d_{\mathcal{H}^{(0)}}(x)/k\Big| \le 2\sqrt{D^{(0)}\log D^{(0)}}.$$

Let $\mathcal{H}^{(1)}$ be a hypergraph satisfying the degree conditions, and $k_1 = 2^{l-1}$,

$$D^{(1)} = 2^{l-1}D^{(0)}/k + 2\sqrt{D^{(0)}\log D^{(0)}}.$$

Then $\mathcal{H}^{(1)}$ is k_1-uniform and

$$D^{(1)} - f(D^{(1)}) \leq d_{\mathcal{H}^{(1)}}(x) \leq D^{(1)},$$

with $e^{k-1}MC\log(1+C)/5 \leq D^{(1)} \leq e^{k-1}MC\log(1+C)/2$. In particular, $D^{(1)}$ is much larger than $k_1^3 C\log(kC)$ as long as M is large enough, and Corollary 3.3 may be applied for the smallest t_1 satisfying $\alpha_{t_1} \leq 2/5$. Thus $D^{(2)} := D_{t_1}$ is between $0.9(2/5)^{k_1-1}D^{(1)}$ and $(2/5)^{k_1-1}D^{(1)}$, or simply

$$(2/5)^{k_1}D^{(1)} \leq D^{(2)} \leq (2/5)^{k_1-1}D^{(1)}.$$

Furthermore, the argument above yields

$$I_{D^{(1)},k_1}(f, g : t_1) = O\Big((C\log(1+C)^{1/(k_1-1)}\Big),$$

and

$$\frac{2|V^{(1)}|I_{D^{(1)},k_1}(f, g : t_1)}{(k_1-1)(D^{(1)})^{1/(k_1-1)}} \leq \frac{2|V^{(1)}|}{k_1-1}.$$

Thus there is a partial cover $\mathcal{D}^{(1)}$ and a hypergraph $\mathcal{H}^{(2)}$ on the uncoverd vertices such that

$$|\mathcal{D}^{(1)}| \leq \frac{3|V^{(1)}|}{k_1} \leq \frac{12|V^{(0)}|}{k}$$

for $V^{(1)} = V^{(0)}$ and $k_1 \geq k/4$, and

$$|V^{(2)}| \leq \frac{2|V^{(1)}|}{5}, \quad D^{(2)} - f(D^{(2)}) \leq d_{\mathcal{H}^{(2)}}(x) \leq D^{(2)}.$$

We iteratively apply the same argument for the parameters $k_i = 2^{l-i}$ and $D^{(i)}$ with

$$(2/5)^{k_{i-1}}D_{i-1} \leq D_i \leq (2/5)^{k_{i-1}-1}D_{i-1}.$$

Since

$$D^{(i)} \geq (2/5)^{2k_1}D^{(1)} \geq (2/5)^{2k_1}e^{k-1}MC\log(1+C)/5 \geq (2e/5)^{k-1}MC\log(1+C)/5$$

there are partial covers $\mathcal{D}^{(i)}$ and hypergraphs $\mathcal{H}^{(i+1)}$ on the uncoverd vertices by $\cup_{j=1}^{i}\mathcal{D}^{(j)}$ such that

$$|\mathcal{D}^{(i)}| \leq 3|V^{(i)}|/k_i$$

and

$$|V^{(i+1)}| \leq \frac{2|V^{(i)}|}{5}, \quad D^{(i+1)} - f(D^{(i+1)}) \leq d_{\mathcal{H}^{(i+1)}}(x) \leq D^{(i+1)}.$$

This procedure stops when $k_i = 4$, in which case

$$|V^{(i)}| \leq (2/5)^{l-3}|V^{(1)}| = o(|V^{(0)}|/k)$$

as $V^{(1)} = V^{(0)}$ and $l \geq \log_2 k - 1$. Therefore, there is a complete cover of size at most

$$\frac{3|V^{(1)}|}{k_1}\sum_{i=0}^{\infty}(4/5)^i + o(|V_0|/k) = O(|V_0|/k).$$

□

Corollary 3.6 *Let $\mathcal{H}$ be a k-uniform hypergraph satisfying (8) and suppose $k > 4$. Then there is a cover of size at most*

$$\frac{|V|}{k}\Big(1+O\Big(\Big(\frac{C\log(1+C)}{D}\Big)^{1/(k-1)}\Big)\Big).$$

In case the degrees of some vertices are so small that (8) is violated, we may still find an economical cover. Suppose a k-uniform hypergraph satisfies (8) except for the vertices in a set B (B for bad), whose degrees are less than $D - f(D)$. For a fixed vertex $x \in B$, consider a hypergraph consisting of k vertex disjoint copies $\mathcal{H}_1, ..., \mathcal{H}_k$ of $\mathcal{H}$ together with an artificial edge containing the k copies of x. Then the degree of x in this hypergraph is one more than the degree in $\mathcal{H}$ and all other degrees remain the same. By this way, we may construct many copies of $\mathcal{H}$ together with artificial edges consisting of copies of vertices in B so that the new hypergraph satisfies (8). Now apply our result to this hypergraph to find an economical cover of the appropriate size and then discard all artificial edges from the cover. Clearly each edge of the partial cover obtained is contained in a unique copy of $\mathcal{H}$. Thus one can find a partial cover of a copy of $\mathcal{H}$ as described in (10), (11), or Corollary 3.6 and it covers all vertices but those in (the copy of) B. As usual, one may choose, for each vertex in B, an edge of (the copy of) $\mathcal{H}$ containing it to extend it to a cover. We summarize these assertions in the theorem below.

Theorem 3.7 *Let $\mathcal{H}$ be a k-uniform hypergraph satisfying (8) except for vertices v in a set B, whose degrees are less than $D - f(D)$. Then*

there is a cover of size at most

$$\mathrm{cov}(\mathcal{H}) = \frac{|\mathrm{V}|}{4}\Big(1+\mathrm{O}\Big(\Big(\frac{(\log \mathrm{D})^4\mathrm{C}}{\mathrm{D}}\Big)^{1/3}\Big)\Big) + |\mathrm{B}|$$

if $k = 4$, and

$$\mathrm{cov}(\mathcal{H}) = \frac{|\mathrm{V}|}{\mathrm{k}}\Big(1+\mathrm{O}\Big(\Big(\frac{\mathrm{C}\log(1+\mathrm{C})}{\mathrm{D}}\Big)^{1/(\mathrm{k}-1)}\Big)\Big) + |\mathrm{B}|$$

if $k > 4$.

If $\mathcal{H}$ is D-regular then B is empty and Theorem 3.7 implies Theorem 1.2. Moreover, if $|B| = o(|V|/k)$, and $k > 4$ with

$$k = o\Big(\log\frac{D}{C\log(1+C)}\Big),$$

then there is a nearly optimal cover of size $(1 + o(1))n/k$.

In case k is so large that the second condition of (8) is violated, we can still derive a bound from Theorem 3.7 by the following procedure. Set

$$k_0 = (1/3)\log(D/(C\log D))$$

and take l such that $lk_0 \leq k < (l+1)k_0$. From each edge, randomly discard $k - lk_0$ vertices and split the remainder into l edges of sizes k_0. Assuming $C = o(D/\log D)$, the same argument used in the proof of Lemma 3.5 would yield a k_0-uniform hypergraph satisfying the degree conditions with respect to some parameter D' with $D/2 \leq D' \leq D$. Thus we have the following corollary.

Corollary 3.8 *Let $\mathcal{H}$ be a k-uniform hypergraph satisfying the first condition of (8) except vertices in a set B, whose degrees are less than $D - f(D)$. If*

$$C = o\Big(\frac{D}{\log D}\Big), \quad and \quad k \geq (1/3)\log(D/C\log D),$$

then there is a cover of size at most

$$O\Big(\frac{|V|}{\log(D/(C\log D))}\Big) + |B|.$$

Theorem 3.7 also implies the following result on matchings.

Theorem 3.9 *Let $\mathcal{H}$ be a k-uniform hypergraph satisfying (8) except vertices in a set B, whose degrees are less than $D - f(D)$. Then there is a matching which covers all vertices except*

$$O\Big(\Big(\frac{(\log D)^4 C}{D}\Big)^{1/3}|V|\Big) + |B| \quad for\ k = 4,$$

$$O\Big(k\Big(\frac{C\log(1+C)}{D}\Big)^{1/(k-1)}|V|\Big) + |B| \quad for\ all\ k > 4.$$

4 Covering random subsets of the projective plane

In this section we describe a simple application of the main result of the previous section. Let $P = PG(2,q)$ denote the finite projective plane of order q. For a real p between 0 and 1, let $A = A_p$ be a random subset of P obtained by choosing, randomly and independently, each point of P with probability p. Let $c(A)$ denote the minimum number of lines of P whose union covers A. Our objective is to estimate the typical size of $c(A_p)$ as a function of p. Since the whole plane can be covered by $q+1$ lines, for every subset A of P, $c(A) \leq q+1$. By a result of Kahn [13] (formulated in his paper in a dual form), there exists an absolute (large) constant b such that if $p \geq b\log q/q$ then almost surely (that is, with probability that tends to 1 as q tends to infinity), $c(A_p) = q+1$. Therefore, in this case we cannot save even a single line over the trivial bound.
If $p = o(q^{-3/2})$ then it is easy to check that almost surely there are only $o(|A_p|)$ points of A_p that lie in lines containing more than 2 points of A_p, meaning that in this range, almost surely $c(A_p) = (1+o(1))|A_p|/2$. Thus if q and q^2p tend to infinity and q^3p^2 tends to 0 then almost surely $c(A_p) = (1+o(1))q^2p/2$.

In the middle range the situation seems more complicated, but we can always determine the typical value of $c(A_p)$ up to a constant factor. Before dealing with the general case, observe that if, say, $p \geq \frac{\log q}{100q}$ then almost surely $c(A_p) = \Theta(q)$. Indeed, $q+1$ lines always suffice, and even if $p = \frac{\log q}{100q}$ then we need at least $\Omega(q)$ points to cover A_p since then almost surely there are at least $\Omega(q\log q)$ points in A_p and by the standard estimates for Binomial distributions (c.f., e.g., [4], Appendix A) no line contains more than $O(\log q)$ of them. Note also that if, say, $p \leq q^{-1.1}$ then almost surely no line contains more than 20 points of A_p and hence in this case almost surely $c(A_p) = \Theta(|A_p|)$.

The following theorem determines the typical value of $c(A_p)$ up to a constant factor in all the other cases.

Theorem 4.1 *There are two absolute positive constants c_1 and c_2 such that for any $p = f/q$ with $q^{-0.1} \leq f \leq \log q/100$, the size $c(A_p)$ of the minimum number of lines covering all points in A_p satisfies, almost surely,*

$$c_1 \frac{qf \log[\frac{\log q}{f}]}{\log q} \leq c(A_p) \leq c_2 \frac{qf \log[\frac{\log q}{f}]}{\log q}.$$

Before proving this theorem, we need the following technical lemma.

Lemma 4.2 *For all sufficiently large q and all positive reals $p = \frac{f}{q}$ where f satisfies $q^{-0.1} \leq f \leq \frac{\log q}{100}$ the following two assertions hold.*
(i) For $s = 4\frac{\log q}{\log[\frac{\log q}{f}]}$,

$$q^2 \binom{q+1}{s} p^s = o(1).$$

(ii) Define

$$F(s) = \binom{q}{s} p^s (1-p)^{q-s},$$

and let t be the largest integer s such that $F(s) \geq \frac{1}{\sqrt{q}}$. Then

$$\frac{1}{3}\left(\frac{\log q}{\log[\frac{\log q}{f}]}\right) \leq t \leq 2\left(\frac{\log q}{\log[\frac{\log q}{f}]}\right). \tag{14}$$

Proof: We assume, whenever this is needed, that q is sufficiently large. To prove (i), note that since $\frac{\log q}{f} \geq 100$, it follows that

$$\log[\frac{\log q}{f}] \leq [\frac{\log q}{f}]^{0.49},$$

say. Therefore, substituting the value of s we conclude that

$$\binom{q+1}{s} p^s = (1+o(1)) \binom{q}{s} (\frac{f}{q})^s \leq [\frac{ef}{s}]^s$$

$$= (\frac{e}{4}\frac{f}{\log q} \log[\frac{\log q}{f}])^s \leq (\frac{f}{\log q})^{0.51s} = \frac{1}{q^{0.51 \cdot 4}} = o(\frac{1}{q^2}),$$

as needed.

To prove (ii) observe that as

$$\frac{F(s+1)}{F(s)} = \frac{(q-s)}{(s+1)} \frac{f}{q} \frac{1}{(1-p)}$$

it follows that the function F is decreasing for all $s \geq 2f$. Define

$$t_1 = \frac{1}{3}\left(\frac{\log q}{\log[\frac{\log q}{f}]}\right)$$

and

$$t_2 = 2\left(\frac{\log q}{\log[\frac{\log q}{f}]}\right).$$

It is easy to check that $t_1 \geq 2f$ (since $\frac{\log q}{f} \geq 100$.) As $F(s)$ is monotone decreasing for $s \geq 2f$ it suffices to check that $F(t_1) \geq \frac{1}{\sqrt{q}}$ and that $F(t_2) < \frac{1}{\sqrt{q}}$ in order to deduce the assertion in (14). To do so note, first, that as $p \geq \frac{\log q}{100q}$ it follows that for every nonnegative s, $(1-p)^{q-s} \geq \frac{1}{q^{0.02}}$.
Substituting the value of t_1 it follows that

$$F(t_1) \geq \frac{1}{q^{0.03}}(\frac{ef}{t_1})^{t_1} = \frac{1}{q^{0.03}}(\frac{ef\log[\frac{\log q}{f}]}{\frac{1}{3}\log q})^{t_1}$$

$$\geq \frac{1}{q^{0.03}}(\frac{f}{\log q})^{t_1} = \frac{1}{q^{0.03}}\frac{1}{q^{1/3}} > \frac{1}{\sqrt{q}},$$

as needed.
Similarly, since $\frac{e}{2}\log M \leq M^{1/2}$ for all $M \geq 100$, and since $\frac{\log q}{f} \geq 100$, it follows that

$$F(t_2) \leq (\frac{ef}{t_2})^{t_2} = (\frac{ef\log[\frac{\log q}{f}]}{2\log q})^{t_2}$$

$$\leq (\frac{f}{\log q})^{t_2/2} = \frac{1}{q} < \frac{1}{\sqrt{q}}.$$

This completes the proof. □

Proof of Theorem 4.1: The lower bound is simple. Almost surely there are $(1+o(1))(q^2+q+1)p = (1+o(1))qf$ points in A, and no line contains s of them provided $q^2\binom{q+1}{s}p^s = o(1)$, since the expected number of lines containing at least s points is at most $(q^2+q+1)\binom{q+1}{s}p^s$. By Lemma 4.2, part (i) this inequality holds for $s = 4\frac{\log q}{\log[\frac{\log q}{f}]}$, implying the desired lower bound.

To prove the upper bound we choose t such that the probability that a fixed set of q points of P contains precisely t members of A_p is roughly $1/\sqrt{q}$. More precisely, let t be the integer defined in Lemma 4.2, part (ii). Put

$$g = F(t) = \binom{q}{t}p^t(1-p)^{q-t}$$

and note that by definition $g \geq \frac{1}{\sqrt{q}}$ and that by Lemma 4.2, part (ii),

$$\frac{1}{3}(\frac{\log q}{\log[\frac{\log q}{f}]}) \leq t \leq 2(\frac{\log q}{\log[\frac{\log q}{f}]}).$$

Define $k = t+1$, and let H be the k-uniform hypergraph whose vertices are the points of A_p and whose edges are all collections of k points of A_p that lie in a line containing precisely k such points.

The hypergraph H is obviously simple. Moreover, the degree of each of its vertices is a Binomial random variable with parameters g and $q+1$, and hence, by our choice of g and the standard estimates for Binomial distributions all degrees are, almost surely, between $(q+1)g + O((qg)^{1/}(\log q)^{1/2})$ and $(q+1)g - O((qg)^{1/2}(\log q)^{1/2})$. Since here $k \leq \log((q+1)g)$ (as $\frac{\log q}{f} \geq 100$), the desired result now follows from the corollary in the previous section. □

5 Separating random sets of points in the plane

A set $\mathcal{L}$ of lines (in $\mathbf{R}^2$) is said to *separate* the points of a set $\mathcal{S}$ if no two points of $\mathcal{S}$ are in the same component of $\mathbf{R}^2 \setminus \mathcal{L}$. Define $X(\mathcal{S})$ to be the minimal number of lines needed to separate the points of $\mathcal{S}$. Recently, Da Silva and Fukuda [7] gave bounds on $X(\mathcal{S})$ when $\mathcal{S}$ consists of n points. In particular, they noted that if $|\mathcal{S}| = n$ then $\frac{\sqrt{8n-7}-1}{2} \leq X(\mathcal{S}) \leq n-1$ and $X\mathcal{S}) = r(\mathcal{S}) - 1$ if $r(\mathcal{S})$, the maximal number of colinear points of $\mathcal{S}$, is greater than $(n+1)/2$.

In this section we shall study $X(\mathcal{S}_n)$ for a set $\mathcal{S}_n$ of n points chosen at random from the square $[0,1]^2$ with the uniform distribution. Our aim is to show that, with high probability, $X(\mathcal{S}_n)$ has order $n^{2/3}/(\log n)^{1/3}$. As usual, we say that an event *holds with high probability* if for some fixed $\epsilon > 0$ the probability of failure is $O(n^{-\epsilon})$.

Theorem 5.1 *With high probability,*

$$X_n = X(\mathcal{S}_n) = \Theta(n^{2/3}/(\log n)^{1/3}). \tag{15}$$

Proof: In the proof below we shall replace $\mathcal{S}_n$ by $\mathcal{P}_n$, a random set of points with Poisson distribution of intensity n in $[0,1]^2$. This change is justified since, with high probability, $n/2 \leq |\mathcal{P}_n| \leq 2n$, so assertion (15) holds for $\mathcal{S}_n$ iff it holds for $\mathcal{P}_n$. In fact, it we shall consider a Poisson process of intensity n in the entire plane, and define $\mathcal{P}_n = \mathcal{P} \cap [0,1]^2$.

(i) First we prove a lower bound on X_n; as we shall see, this is rather easy. Set

$$k = n^{2/3}/(\log n)^{1/3},$$

and take a $k \times k$ grid in $[0,1]^n$, dividing it into $1/k \times 1/k$ squares called *cells* so that altogether we have k^2 cells. For $j \geq 0$, a j-cell is a cell containing precisely j points of $\mathcal{P}_n$. For each fixed cell, the number of points in the cell has Poisson distribution with mean n/k^2, so the probability p that a fixed cell is a 2-cell is

$$p = \frac{n^2}{2k^4} e^{-n/k^2} \leq \frac{n^2}{2k^4}.$$

Since $n/k^2 = o(1)$, we have $p \sim n^2/2k^4$. In particular, if n is large enough,

$$\frac{2n^2}{5k^4} \leq p \leq \frac{n^2}{2k^4}.$$

The number of 2-cells has binomial distribution $\mathrm{Binom}(k^2, p)$ with mean $k^2 p \geq 2n^2/5k^2$. Hence, with probability $1 + o(e^{-n^{1/2}})$, there are at least $n^2/3k^2$ 2-cells.

In order to prove a lower bound for X_n, all we shall use is that in a separating system of lines every 2-cell is intersected by a line. Fix a line ℓ. Since ℓ intersects at most $2k-1 < 2k$ cells, the probability that ℓ intersects at least d 2-cells is at most

$$\binom{2k}{d} p^d < (\frac{2ekn^2}{d2k^4})^d = (\frac{e \log n}{d})^d.$$

Hence, the probability that a fixed line intersects at least $d_0 = e^2 \log n$ 2-cells is less than $n^{-e^2} < n^{-7}$.

For a line ℓ, let $C(\ell)$ be the set of cells intersected by ℓ. Note that, crudely, for $k \geq 3$ there are at most

$$2(\binom{(k+1)}{2})^2 < k^4$$

sets $C(\ell)$, since $C(\ell)$ can always be given by a line ℓ that goes 'slightly' above or below two of the $(k+1)^2$ 'lattice points' of our $k \times k$ grid, and these two lattice points are neither in the same row nor in the same column. Hence the probability that *some* line meets at least d_0 2-cells is at most

$$k^4 n^{-7} = o(n^{-4}).$$

Therefore, with probability $1 + o(n^{-4})$, there are at least $n^2/3k^2$ 2-cells *and* every line intersects at most d_0 of these 2-cells. Consequently, with probability $1 + o(n^{-4})$, we need at least

$$(n^2/3k^2)/d_0 > \frac{n^{2/3}}{22(\log n)^{1/3}}$$

lines to separate $\mathcal{P}_n$

(ii) Now we turn to the real content of the theorem, the upper bound in equation (15). It will be convenient to take a Poisson process of intensity n in the entire plane and define $\mathcal{P}_n = \mathcal{P} \cap [0,1]^2$. Our strategy is simple. We use a grid to separate points 'far apart', about $\log\log n$ sets of lines to separate pairs of points at 'medium' distance, and one line each for every pair of points at 'small' distance.

Dealing with 'large' and 'small' distances is rather easy, so we shall do it first. Set $\bar{h} = 4n^{2/3}/(\log n)^{1/3}$, and take the $2(\bar{h} - 1) < 8n^{2/3}/(\log n)^{1/3}$ internal lines of this grid. These lines separate all pairs of points in $[0,1]^2$ that are 'far apart', i.e., at distance at least $\sqrt{2}/\bar{h} < \frac{1}{2}(\log n)^{1/3} n^{-2/3}$.

Before we turn to separating points that are close to each other, we consider sets of grids. A *batch* of h-grids is a set of four grids, each with $1/h \times 1/h$ cells: the first is an $h \times h$ grid on $[0,1]^2$, the second is an $(h+1) \times ((h+1)$ grid with lower left vertex at $(-1/2h, -1/2h)$, and the other two are $h \times (h+1)$ and $(h+1) \times h$ grids with lower left vertices at $(0, -1/2h)$ and $(-1/2h, 0)$. Thus all four grids cover $[0,1]^2$ and, given their lower left vertices, are minimal with respect to that. With slight inaccuracy, we shall take each of the four grids of a batch to be an $h \times h$ grid.

All we shall need about a batch of the h-grids is that if $x, y \in [0,1]^2$ with $d(x,y) \leq \frac{1}{2h}$ then x and y belong to the same cell in one of the four grids of the batch.

To separate the pairs of points of $\mathcal{P}_n$ at 'medium' and 'small' distances, for each t with $0 \leq t \leq u = \frac{1}{2\log 2} \log\log n$ we shall make use of the batch $\mathcal{B}_t$ of h_t-grids, where

$$h_t = 2^t n^{2/3}/(\log n)^{1/3} = n^{2/3}/c_t.$$

Thus $h_u = n^{2/3}(\log n)^{1/6}$. Needless to say, we shall not use the lines of these grids, since otherwise we would use many more than $O(n^{2/3}/(\log n)^{1/3})$ lines, but we shall use them to identify the pairs of the points we wish to separate at each stage.

To separate points that are close, we consider the grids of $\mathcal{B}_u$. The probability that a cell of an h_u-grid is $2, 3, \ldots,$ or $7-$ cell is

$$(1 + o(1))\frac{n^2}{2h_u^4} \leq \frac{n^2}{h_u^4},$$

so the expected number of such cells is at most $n^2/h_u^2 \le n^{2/3}/(\log n)^{1/3}$. Hence, with high probability, there are at most $2n^{2/3}/(\log n)^{1/3}$ such cells. The points in these cells can be separated by $12n^{2/3}/(\log n)^{1/3}$ lines. Furthermore, the expected number of i-cells with $i \ge 8$ is

$$(1+o(1))\frac{n^8}{8!h_u^{14}} \le n^{-4/3},$$

so with probability $1+o(1/n)$ neither of the four grids in $\mathcal{B}_u$ contains an i-cell with $i \ge 8$. Since every pair of points $\{x,y\}$ in $\mathcal{P}_n$ with $d(x,y) \le 1/2h_u$ is contained in the same cell of one of the grids in $\mathcal{B}_u$, we see that with probability $1+O(1/n)$ all such pairs can be separated by at most $12n^{2/3}/(\log n)^{1/3}$ lines.

With this, we have arrived at the main part of the proof: we shall show that with high probability, all pairs $\{x,y\} \subset \mathcal{P}_n$ with

$$\frac{1}{2h_u} < d(x,y) < \frac{1}{2}(\log n)^{1/3}/n^{2/3} = \frac{1}{2h_0} \tag{16}$$

can be separated by $O(n^{2/3}\log n)$ lines.

Set $c_i = (\log n)^{1/3}2^{-i}$, so that $h_i = n^{2/3}/c_i$, $0 \le i \le u$. We shall treat the pairs $\{x,y\} \subset \mathcal{P}_n$ with

$$\frac{1}{4h_i} \le d(x,y) \le \frac{1}{2h_i}$$

separately, using the batch $\mathcal{B}_i$. We know that every such pair $\{x,y\}$ is contained in some cell of a grid in $\mathcal{B}_i$, so it suffices to prove that, with high probability, there are

$$O\Big(\frac{n^{2/3}}{2^i(\log n)^{1/3}}\Big) \tag{17}$$

lines that separate all pairs of $\{x,y\} \subset \mathcal{P}_n$ that are contained in some cell of a grid in $\mathcal{B}_i$ and satisfy $d(x,y) \ge 1/4h_i$. Indeed, equation (17) implies that, with high probability, all pairs $\{x,y\}$ satisfying equation (16) can be separated by

$$O\Big(\sum_{i=0}^{u}\frac{n^{2/3}}{2^i(\log n)^{1/3}}\Big) = O\Big(n^{2/3}/(\log n)^{1/3}\Big)$$

lines.

To prove equation (17), set $c = c_i$, and $h = h_i = n^{2/3}/c$ so that $(\log n)^{-1/3} \le c \le (\log n)^{1/3}$ and $n^{2/3}/(\log n)^{1/3} \le h \le n^{2/3}(\log n)^{1/6}$, and let L be a fixed h-grid. Call a cell of L *full* if it contains at least two points of $\mathcal{P}_n$, and write p for the probability that a cell is full, so that

$$p = e^{-n/h^2}\sum_{j=2}^{\infty}\frac{(n/h^2)^2}{j!} = (1+o(1))\frac{n^2}{2h^4}. \tag{18}$$

Our aim is to show that, with high probability, there is a family $\mathcal{F}_h$ of $O(n^{2/3}/2^i(\log n)^{1/3})$ 'almost horizontal' lines such that each full cell is met by one of these lines. Here 'almost horizontal' means that for some $\beta-\alpha > 0$ each line has slope at most $h^{\alpha-\beta}$. Replace each line $\ell \in \mathcal{F}_h$ by a family of 17 lines, say, obtained from ℓ by translating it in the vertical direction by $j/8h, j = 0, \pm 1, \ldots, \pm 8$. The new family $\mathcal{F}_h^*$ of $O(n^{2/3}/2^i(\log n)^{1/3})$ lines is such that if $x = (x_1, x_2)$ and $y = (y_1, y_2)$ are two points of $\mathcal{P}_n$ in the same cell of the grid with $|x_2 - y_2| \geq \frac{1}{4\sqrt{2}h}$ then some line $\ell \in \mathcal{F}_h^*$ separates x from y. By symmetry, there is a similar family $\mathcal{G}_h^*$ of 'almost vertical' lines. Clearly, $\mathcal{F}_h^* \cup \mathcal{G}_h^*$ separates all points $x, y \in \mathcal{P}_n$ in the same cell of L provided $d(x,y) \geq \frac{1}{4h}$.

We have further narrowed our problem: we have to show that $\mathcal{F}_h$ exists with high probability. To this end, let α, β, γ and δ be constants with

$$0 < \gamma < \alpha < \beta < \alpha + \beta < \delta < 1.$$

For each i, $1 \leq i \leq h^\alpha$, set $e_i = \lfloor ih^{1-\beta} \rfloor$ and $\varepsilon_i = \frac{e_i}{h-e_i} \sim ih^{-\beta}$. Let $\mathcal{H}_i$ be the family of h lines with slope ε_i, such that the jth line goes through the point $(1 - j/h + \eta, (j-1)/h)$, where $0 < \eta < 1/h$ is chosen so that none of the h lines goes through a point of L.

Note that the jth line crosses the line $y = 0$ at $x = a_j = 1 - j/h + \eta - (j-1)/h\varepsilon_i$ and the line $y = 1$ at $x = b_j = 1 - j/h + \eta + (h - j + 1)/h\varepsilon_i)$. Our choice of ε_i implies that $a_j = b_{h-e_i+j}$, precisely the first e_i lines cross the bottom of the grid L, and precisely the e_i lines cross the top of the grid L. For $1 \leq j \leq e_i$, we identify the jth line with the $(h - e_i + j)$th line, so that $\mathcal{H}_i$ consists of $h - e_i$ lines. Equivalently, we take the y coordinates of the lines of H_i modulo 1.

For each line $\ell \in \mathcal{H}_i$, let $C(\ell)$ be the set of cells of the grid L intersected by ℓ. We call $C(\ell)$ a *combinatorial line*. Let $\mathcal{C}_i$ be the set of combinatorial lines obtained from the lines of $\mathcal{H}_i$ and set $\mathcal{C} = \bigcup_{i=1}^{h^\alpha} \mathcal{C}_i$. By construction, every line in $\mathcal{H}_i$ crosses either $\lfloor \varepsilon_i h \rfloor$ or $\lceil \varepsilon_i h \rceil$ horizontal division lines of the grid L, so it crosses $h + \lfloor \varepsilon_i h \rfloor$ or $h + \lceil \varepsilon_i h \rceil$ cells of L. Also, every cell of L is in precisely one combinatorial line of $\mathcal{C}_i$. Thus the combinatorial lines of $\mathcal{C}_i$ partition the h^2 cells of the L into $h - e_i$ sets of as equal sizes as possible. Crudely, each combinatorial line has at least h and at most $h + 2h^{\alpha-\beta}$ cells.

In addition to the 'equipartition' property of $\mathcal{C}_i$, we need the following simple consequence of the fact that for $i \neq j$ the slopes ε_i and e_j differ by at least about $h^{-\beta}$. If $i \neq j$, $C_i \in \mathcal{C}_i$ and $C_j \in \mathcal{C}_j$, then no two cells of $C_i \cap C_j$ are horizontally separated, where we call two cells of L *horizontally separated* if either no combinatorial line contains both of them, or their x coordinates differ by at least $2h^{\beta-1}$.

Let us turn the question of finding a suitable set of lines $\mathcal{F}_h$ into a question about hypergraph covering. Let $0 < \gamma < \alpha$ and let m be the maximal integer such that

$$g = \binom{h}{m} p^m (1-p)^{h-m} \geq h^{-\gamma}, \tag{19}$$

where p is as in (18). Clearly, g is the probability that precisely m of h given cells are full. Standard estimates show that

$$m = \Theta(\frac{\log n}{t+1}) \tag{20}$$

Set $k = m + 1$, and define a k-uniform hypergraph H as follows. The vertex set of H is the set V of the full cells of L. A k-subset of V is an edge of H if it is of the form $C \cap V$ for some combinatorial line $C \in \mathcal{C}$ and it consists of horizontally separated cells. The construction of H implies that it is a

k-uniform simple hypergraph. Furthermore, $|V|$ has binomial distribution with parameters h^2 and p so, with high probability, $|V| = (1+o(1))h^2p = (\frac{1}{2}+o(1))n^2/h^2 = (\frac{1}{2}+o(1))c^3h$.

To complete the proof of our theorem, it suffices to show that, with high probability, H has a cover with

$$O\Big(\frac{n^{2/3}}{2^i(\log n)^{1/3}}\Big) \tag{21}$$

edges.

Indeed, if H has a cover with q edges then the set of full cells of the grid L is the union of some q combinatorial lines, and so there is a family $\mathcal{F}_h$ of at most $2q$ 'almost horizontal' lines such that every full cell is intersected by at least one of these lines.

To prove the bound (21) we shall show that, with high probability, H is an almost regular simple hypergraph except for a few vertices of low degree, and then apply our hypergraph covering theorem.

Call a full cell σ *horizontally isolated* if it is horizontally separated from every other full cell. Clearly, these are at most

$$2(2h^\beta)(2h^\beta h^{\alpha-\beta}) = 8h^{\alpha+\beta}$$

cells that are not horizontally separated from σ. Hence, writing W for the set of cells that are full but not horizontally isolated, we see that

$$\mathbf{E}(|W|) \le h^2p^2 8h^{\alpha+\beta} = O((\log n)^2 h^{\alpha+\beta}).$$

Therefore, with high probability, $|W| < h^\delta$. The sets W are the exceptional vertices of H.

Since $m < (\log n)^2$, for $h' = h + O(h^{\alpha-\beta})$ we have

$$\binom{h'}{m} p^m (1-p)^{h'-m} = (1+o(1))g.$$

Furthermore, if we select m numbers from $1, 2, \dots, h'$ then almost every selection is such that any two selected numbers differ by more than $3h^\beta$. Therefore, conditional on a cell σ being full and horizontally isolated, the probability that a combinatorial line through σ gives an edge of H is $(1+o(1))g$. Consequently, the conditional expectation of the degree of σ, $\mathbf{E}'(d(\sigma))$, is $(1+o(1))gh^\alpha \ge (1+o(1))h^{\alpha-\gamma}$.

We claim that, with high probability, every full and horizontally isolated cell σ is such that

$$|d(\sigma) - \mathbf{E}'(d(\sigma))| \le 4(gh^\alpha \log h)^{1/2}. \tag{22}$$

To see (22), note that, conditional on a fixed cell σ being full and horizontally isolated, $d(\sigma)$ is the sum of h^α independent Bernoulli random variables, with probabilities $g_1, g_2, \dots, g_{h^\alpha}$, where $g_i = (1+o(1))g$ for each i. Therefore the probability that (20) fails is at most h^{-7}, and so the probability that it fails for some cell is at most h^{-5}.

A weak form of (22) is that, with high probability, every full and horizontally isolated cell has degree $(1+o(1))gh^\alpha$. In fact, with high probability, every vertex of H has degree at most $(1+o(1))gh^\alpha$ since, for every full cell σ, if we fix the distribution of full cells horizontally separated from σ then the degree of σ is maximal if σ is horizontally isolated.

Let us summarize what we have learned about H: with high probability, H is a k-uniform simple hypergraph with $(\frac{1}{2}+o(1))c^3h$ vertices and maximal degree $(1+o(1))gh^{\alpha} \geq (1+o(1))h^{\alpha-\gamma}$ such that all but at most h^{δ} vertices have degree $(1+o(1))gh^{\alpha}$. Since $k = O(\log h^{\alpha-\gamma})$ and $h^{\delta} < c^3h/k$, our hypergraph cover theorem implies that such an H has a cover with $O(c^3h/k)$ edges. Recalling (20) and $c = c_i = (\log n)^{1/3}2^{-i}$, we see that, with high probability, H has a cover with

$$O\Big(\frac{(\log n)2^{-3i}n^{2/3}2^i(i+1)}{(\log n)^{4/3}}\Big) = O\Big(\frac{n^{2/3}(i+1)}{2^{2i}(\log n)^{1/3}}\Big),$$

implying (21) and so completing the proof.

□

The method above can be used to solve the higher dimensional analogue of Theorem 5.1, but the technical difficulties involved are considerably more unpleasant.

6 Concluding remarks

- It seems plausible that for every k and D bigger than 1, the minimum size of a cover of a simple D-regular k-uniform hypergraph on n vertices is

  $$O(\ Min\{\frac{n}{k}\log(\frac{k}{\log D}+2),\ \frac{n}{k}D\}),$$

  and that this is tight, up to a constant factor, for all admissible values of k and D. Our methods here suffice to show this is the case for all $k \leq O(\log D)$ as well as for all $k \geq (\log D)^{1+\Omega(1)}$.

- It will be interesting to determine the maximum possible $k = k(D)$ for which every simple D-regular k-uniform hypergraph on n vertices contains a cover of nearly optimal size, that is, of size $(1+o(1))\frac{n}{k}$, where the $o(1)$-term tends to 0 as D tends to infinity. By the results in Sections 2 and 3 this holds when $k(D) = o(\log D)$ and does not hold when $k(D) > (1+\Omega(1))e\log D$. It is not difficult to see that if a nearly optimal cover always exists for k and D, and if $k' < k$, then a nearly optimal cover always exists for k' and D as well.

- It seems that the arguments in Section 5 can be extended to higher fixed dimensions, using the same technique. Thus, for every fixed d, the minimum number of hyperplanes needed to separate all points in a randomly chosen set of n points in the unit cube in R^d is, almost surely, $\Theta(\frac{n^{2/(d+1)}}{(\log n)^{1/(d+1)}})$.

References

[1] M. Ajtai and J. Komlós and E. Szemeréd, *A dense infinite Sidon sequence*, Europ. J. Combinatorics 2 (1981), 1-11.

[2] N. Alon, Transversal numbers of uniform hypergraphs, Graphs and Combinatorics 6 (1990), 1-4.

[3] N. Alon, J. H. Kim and J. Spencer, *Nearly perfect matchings in regular simple hypergraphs*, Israel J. Math. 100 (1997), 171–187.

[4] N. Alon and J. H. Spencer, *The Probabilistic Method, Second Edition*, Wiley, New York, 2000.

[5] K. Azuma, *Weighted sums of certain dependent random variables,* Tokuku Math. J. **19** (1967), 357-367.

[6] B. Bollobás, A probabilistic proof of an asymptotic formula for the number of labelled regular graphs, European J. Combin. 1 (1980), 311–316.

[7] I.P.F. Da Silva and K. Fukuda, Isolating lines in the plane, *J. Geometry* **62** (1998), 48–65.

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

[9] P. Frankl and V. Rödl, *Near perfect coverings in graphs and hypergraphs*, Europ. J. Combinatorics 6 (1985), 317-326.

[10] D. A. Grable, *More-than-nearly perfect packings and partial designs*, Preprint, 1995.

[11] W. Hoeffding, *Probability inequalities for sums of bounded random variables*, J. Amer. Stat. Assoc. 27 (1963), 13-30.

[12] A. Johansson, *An improved upper bound on the choice number for triangle free graphs*. Preprint.

[13] J. Kahn, On a problem of Erdős and Lovász: random lines in a projective plane, Combinatorica 12 (1992), 417–423.

[14] J. Kahn, *Coloring Nearly-Disjoint Hypergraphs with $n + o(n)$ Colors*, J. Combinatorial Theory, Ser. A 59 (1991), 31-39.

[15] J. Kahn, *Asymptotically good list-colorings*, J. Combinatorial Theory, Ser. A 73 (1996), 1–59.

[16] J.-H. Kim, *On Brooks' theorem for sparse graphs*, Combinatorics, Probability & Computing 4 (1995), 97-132.

[17] J.-H. Kim, The Ramsey number $R(3, t)$ has order of Magnitude $t^2/\log t$, *Random Structures and Algorithms* 7 (1995), 173-207.

[18] J.-H. Kim, Nearly Optimal Partial Steiner Systems. Preprint.

[19] J.-H. Kim and V. H. Vu, *Small complete arcs on finite projective planes*, Submitted.

[20] J.-H. Kim and V. H. Vu, *Concentration of multivariate polynomials and its applications*, Combinatorica 20 (2000), 417–434.

[21] A. Kostochka and V. Rödl, *Partial Steiner systems and matchings in hypergraphs*, Random Structures and Algorithms 13 (1998), 235-240.

[22] M. Molloy and B. Reed, *A bound on the total chromatic number*, Combinatorica 18 (1998), 241-280.

[23] N. Pippenger, unpublished.

[24] N. Pippenger and J. H. Spencer, *Asymptotic behaviour of the chromatic index for hypergraphs*, J. Combinatorial Theory, Ser. A 51 (1989), 24-42.

[25] V. Rödl, *On a packing and covering problem*, Europ. J. Combinatorics 6 (1985), 69-78.

[26] M. Talagrand, Concentration of Measures and Isopermetric Inequalites in product spaces, *Publications Mathematiques de l'I.H.E.S.,* 81 (1996), 73-205.

[27] V. H. Vu, On nearly perfect matchings of hypergraphs: higher codegrees do help, *Random Structures and Algorithms* 17 (2000), 29-63.

[28] V. H. Vu, *On some simple degree conditions that guarantee the upper bound on the chromatic (choice) number of random graphs.* J. Graph Theory 31 (1999), 201–226.

[29] V. H. Vu, A new upper bound on the list chrmatic number of locally sparse graphs, *to appear in CPC.*