Bounds for Optimal Coverings

Carlos Gustavo T. de A. Moreira [a] and Yoshiharu Kohayakawa [b]

[a] *IMPA – Estrada Dona Castorina 110, 22460–320 Rio de Janeiro, RJ, Brazil – e-mail: gugu@impa.br*

[b] *Instituto de Matemática e Estatística, USP – Rua do Matão 1010, 05508–090 São Paulo, SP, Brazil – e-mail: yoshi@ime.usp.br*

Abstract

We give bounds for optimal coverings of finite sets by elements of regular families of subsets, and show that both upper and lower bounds are asymptotically sharp for some families of examples.

Key words: Coverings, set systems, hypergraphs, the probabilistic method

1 Introduction

We study coverings of finite sets by subsets belonging to a regular family. By an (a,b)-*regular* family of subsets of a finite set F, or an (a,b)-*regular hypergraph on* F, we mean a family $\mathcal{C}$ for which there are integers a and b such that for each $C \in \mathcal{C}$ we have $|C| = a$, and, for each $x \in F$, we have $\deg(x) = b$, where as usual $\deg(x) = \Big|\{C \in \mathcal{C} \mid x \in C\}\Big|$. Thus (a,b)-regular families are what are sometimes called a-*uniform*, b-*regular* hypergraphs. The problem of constructing small covers for such hypergraphs and of estimating the minimum possible size for such covers are common, and appear in many contexts, as in the study of covering codes (see [3]), in particular, in football pool problems (see [6]), in the study of vertex covers of graphs, and others. Somewhat surprisingly, some sharp results are known for problems of this kind, see, e.g., [1], [4, Theorem 8.11], [5], and [9]. In this paper, we prove another result that

* Partially supported by MCT/FINEP/CNPq (ProNEx Projects 416/96 and 107/97), FAPESP (Proc. 96/04505–2), CNPq (Proc. 300647/95–6, 300334/93–1, and 468516/2000–0), and by FAPERJ.

**The authors gratefully acknowledge the careful work of the referees.

shows that one may require strong regularity conditions and still obtain hypergraphs that behave asymptotically ‘badly’ with respect to the size of covers that they admit.

Let us now introduce the extremal parameter that we are interested in. If $\mathcal{C}$ is a hypergraph on F, we let $\alpha(F,\mathcal{C})$ be the minimal integer r for which there exist $C_1,\dots,C_r\in\mathcal{C}$ such that $\bigcup_{1\le j\le r}C_j=F$. We are concerned with estimating the extremal values of $\alpha(F,\mathcal{C})$.

This paper is organized as follows. In Sections 2 and 3, we prove our upper and lower bounds for $\alpha(F,\mathcal{C})$. Our proof for the existence of systems $\mathcal{C}$ with large covering number $\alpha(F,\mathcal{C})$ is probabilistic; in Section 3.2, we briefly discuss a construction of Alon [1] based on finite fields and character sums. In Section 4 we mention the dual problem of estimating the packing number of hypergraphs. We conclude with some remarks and an open problem.

2 Upper bounds

Our first result is a slight improvement of classical results by Johnson, Stein, and Lovász (see [7], [8], and [10]). Let us say that a family $\mathcal{C}$ of subsets of a set F is (a,b)-*semiregular* if $|C|\le a$ for all $C\in\mathcal{C}$ and $\deg(x)\ge b$ for all $x\in F$.

Proposition 1 *Let a and b be positive integers and suppose $\mathcal{C}$ is an (a,b)-semiregular family on an n-element set F. Let $m=|\mathcal{C}|$. Then, for any positive integer ℓ',*

$$\frac{n}{a}\le\alpha(F,\mathcal{C})\le\frac{\ln(m\ell'/bn)}{\ln(1-b/m)}+\frac{m}{b}\sum_{1\le j\le\ell'}\frac{1}{j}. \tag{1}$$

In particular, we have $\alpha(F,\mathcal{C})\le f(a,b,m,n)$, where we let $\ell=\lceil bn/m\rceil$ and

$$f(a,b,m,n)=\begin{cases}\dfrac{n}{\ell}+\dfrac{m}{b}\displaystyle\sum_{j=2}^{\ell}\frac{1}{j} & \text{if } b\le m/\sqrt{n}\\[2ex] \dfrac{m}{b}\left(\ln\left(\dfrac{bn}{m}\right)+\gamma\right)-\dfrac{1}{2}\ln\left(\dfrac{b^2n}{m^2}\right)+\dfrac{1}{2} & \text{if } b>m/\sqrt{n}.\end{cases} \tag{2}$$

Moreover, if $b\le m/\sqrt{n}$, we have

$$f(a,b,m,n)\le\frac{n}{a}+\frac{m}{b}\sum_{2\le j\le a}\frac{1}{j}<\frac{m}{b}(\ln(a+1)+\gamma), \tag{3}$$

where $\gamma=0.5772156649\ldots$ is Euler’s constant.

We prove Proposition 1 in Section 2.1 below.

Corollary 2 *Let F be an n-element set and $\mathcal{C}$ an (a,b)-regular hypergraph on F. Let also $\ell = \min\{a, \lceil n/a \rceil\}$. Then*

$$\left\lceil \frac{n}{a} \right\rceil \le \alpha(F,\mathcal{C}) \le \frac{\ln(a/\ell)}{\ln \dfrac{1}{1-a/n}} + \frac{n}{a}\sum_{j=1}^{\ell}\frac{1}{j}$$
$$< \frac{n}{a}\Big(\ln(a+1)+\gamma\Big) - \frac{1}{2}\ln\frac{a}{\ell} + \frac{1}{2}. \qquad (4)$$

Our main result, given in Section 3 (see Proposition 6), shows that the estimates for $\alpha(F,\mathcal{C})$ above are asymptotically sharp for certain regular families of subsets of finite sets.

2.1 Proof of Proposition 1

We state and prove two auxiliary lemmas first. For the remainder of this section, we fix an (a,b)-semiregular family $\mathcal{C}$ on a set F, where a and b are positive integers. We also let $n = |F|$ and $m = |\mathcal{C}|$. Our first lemma follows from a well known double counting argument.

Lemma 3 *We have $am \ge bn$. Moreover, given any $A \subset F$, there is $C \in \mathcal{C}$ such that $|C \cap A| \ge (b/m)|A|$.*

PROOF. Indeed, for any $B \subset F$, we have

$$ma \ge \sum_{C\in\mathcal{C}} |C\cap B| = \sum_{C\in\mathcal{C}}\sum_{x\in B}\chi_C(x) = \sum_{x\in B}\sum_{C\in\mathcal{C}}\chi_C(x) = \sum_{x\in B}\deg(x) \ge b|B|,$$

where, as usual, χ_C is the characteristic function for the set C, that is, $\chi_C(x) = 1$ if $x \in C$ and $\chi_C(x) = 0$ otherwise. Taking $B = F$, we obtain $am \ge nb$, which proves the first inequality in our lemma. Taking $B = A$, we deduce that $\sum_{C\in\mathcal{C}}|C\cap A| \ge b|A|$, which implies that there is a set $C \in \mathcal{C}$ for which $|C \cap A| \ge (b/|\mathcal{C}|)|A|$, as required. $\square$

An immediate corollary to Lemma 3 is the following.

Corollary 4 *For each positive integer k, there are $C_1, \ldots, C_k \in \mathcal{C}$ such that*

$$\left| F \setminus \bigcup_{1\le i\le k} C_i \right| \le n\left(1 - \frac{b}{m}\right)^k. \qquad (5)$$

A finer corollary to Lemma 3 is as follows.

Corollary 5 *It is possible to cover any $A \subset F$ with at most*

$$\frac{r}{\ell} + \frac{m}{b} \sum_{2 \le j \le \ell} \frac{1}{j} \le \frac{m}{b} \sum_{1 \le j \le \ell} \frac{1}{j} \tag{6}$$

members of $\mathcal{C}$, where $\ell = \lceil br/m \rceil$ and $r = |A|$.

PROOF. We can cover A by $C_1, C_2, \ldots \in \mathcal{C}$ so that, for each i, the cardinality of $C_i \cap (A \setminus \bigcup_{1 \le j < i} C_j)$ is as large as possible, given $C_1, \ldots, C_{i-1}$. For $1 \le s \le a$, let k_s be the number of sets C_i in this covering such that $|C_i \cap (A \setminus \bigcup_{1 \le j < i} C_j)| = s$. A little thought shows that Lemma 3 implies that

$$k_1 \le \frac{m}{b},\ k_1 + 2k_2 \le \frac{2m}{b},\ \ldots\ ,\ k_1 + 2k_2 + \cdots + sk_s \le \frac{ms}{b} \tag{7}$$

for all s. Moreover, clearly,

$$k_1 + 2k_2 + \cdots + sk_s \le r \tag{8}$$

for all s as well. Of course

$$k_1 + 2k_2 + \cdots + ak_a = r. \tag{9}$$

Note also that $a \ge \ell$. From (7)–(9) it thus follows that

$$\begin{aligned} & k_1 + \cdots + k_a \\ &\quad = \frac{1}{a}(k_1 + 2k_2 + \cdots + ak_a) + \sum_{1 \le j < a} \frac{1}{j(j+1)}(k_1 + 2k_2 + \cdots + jk_j) \\ &\quad\quad \le \frac{r}{a} + \frac{m}{b} \sum_{1 \le j < \ell} \frac{1}{j+1} + r \sum_{\ell \le j < a} \frac{1}{j(j+1)} \\ &\quad\quad\quad = \frac{r}{\ell} + \frac{m}{b} \sum_{1 \le j < \ell} \frac{1}{j+1} \le \frac{m}{b} \sum_{1 \le j \le \ell} \frac{1}{j}, \end{aligned}$$

as required. □

We may finally prove Proposition 1.

PROOF. (Proof of Proposition 1) In the case $b \le m/\sqrt{n}$, the result in (2) follows from Corollary 5 by taking $r = n$. The first inequality in (3) follows from

$$\frac{n}{\ell} = \frac{n}{a} + \sum_{\ell \le j < a} \frac{n}{j(j+1)} \le \frac{n}{a} + \frac{m}{b} \sum_{\ell \le j < a} \frac{\ell}{j(j+1)} \le \frac{n}{a} + \frac{m}{b} \sum_{\ell \le j < a} \frac{1}{j+1},$$

and hence

$$\frac{n}{\ell} + \frac{m}{b} \sum_{2\le j\le \ell} \frac{1}{j} \le \frac{n}{a} + \frac{m}{b} \sum_{2\le j\le a} \frac{1}{j}.$$

The second inequality in (3) is clear.

In general, given a positive integer ℓ', we claim that the upper bound in (1) holds. To prove this claim, take $r' = m\ell'/b$, and let

$$\frac{\ln(m\ell'/bn)}{\ln(1-b/m)} = p + c, \tag{10}$$

where p is an integer and $0 \le c < 1$. Let $k = p+1$, and set

$$r = \left\lfloor n\left(1-\frac{b}{m}\right)^{k}\right\rfloor \le r'. \tag{11}$$

We now apply Corollary 4 with the above value of k; this leaves us with an uncovered set A of cardinality at most r. We then apply Corollary 5 with $\ell = \lceil br/m\rceil \le \lceil br'/m\rceil = \ell'$ to cover A. Using (10), (11), and the fact that the left-hand side of (6) is increasing in ℓ for $\ell \ge br/m$, we deduce that this application of Corollary 5 uses at most

$$\begin{aligned}
\frac{r}{\ell} + \frac{m}{b}\sum_{2\le j\le \ell}\frac{1}{j} \le \frac{r}{\ell'} + \frac{m}{b}\sum_{2\le j\le \ell'}\frac{1}{j} &\le \frac{1}{\ell'}\, n\left(1-\frac{b}{m}\right)^{p+c+1-c} + \frac{m}{b}\sum_{2\le j\le \ell'}\frac{1}{j} \\
&\le \frac{n}{\ell'}\left(\frac{m\ell'}{bn}\right)\left(1-\frac{b}{m}\right)^{1-c} + \frac{m}{b}\sum_{2\le j\le \ell'}\frac{1}{j} \\
&= \frac{m}{b}\left(1-\frac{b}{m}\right)^{1-c} + \frac{m}{b}\sum_{2\le j\le \ell'}\frac{1}{j}
\end{aligned} \tag{12}$$

members of $\mathcal{C}$. Therefore, in this covering we have used in total at most

$$\begin{aligned}
&(p+1) + \frac{m}{b}\left(1-\frac{b}{m}\right)^{1-c} + \frac{m}{b}\sum_{2\le j\le \ell'}\frac{1}{j} \\
&\qquad = \frac{\ln(m\ell'/bn)}{\ln(1-b/m)} + (1-c) + \frac{m}{b}\left(1-\frac{b}{m}\right)^{1-c} + \frac{m}{b}\sum_{2\le j\le \ell'}\frac{1}{j}
\end{aligned} \tag{13}$$

members of $\mathcal{C}$. The function $\phi(c) = 1 - c + (m/b)(1-b/m)^{1-c}$ is convex, as may be seen by computing its second derivative, and hence its maximum is $\phi(0) = \phi(1) = m/b$. Therefore (13) is bounded by

$$\frac{\ln(m\ell'/bn)}{\ln\left(1-b/m\right)} + \frac{m}{b}\sum_{1\le j\le \ell'}\frac{1}{j},$$

as required.

In order to prove the inequality $\alpha(F, \mathcal{C}) \leq f(a, b, m, n)$ in the case $b > m/\sqrt{n}$, we just take $\ell' = \lfloor m/b \rfloor$ and apply the above inequality.

Some calculations complete the proof. The two main ingredients in these calculations are the inequalities

$$\sum_{1 \leq r \leq k} \frac{1}{r} - \ln k - \gamma < \frac{1}{2k} - \frac{1}{12k(k+1)},$$

which is valid for every positive integer k, and

$$-\frac{1}{\ln(1-x)} < \frac{1}{x} - \frac{1}{2},$$

valid for all $0 < x < 1$. $\square$

2.2 Proof of Corollary 2

Since $\mathcal{C}$ is an (a, b)-regular family, we have $ma = nb$. We take $\ell' = \ell$ in Proposition 1. Note that, then, the right-hand side of (1) becomes

$$\frac{\ln(a/\ell)}{\ln \dfrac{1}{1 - a/n}} + \frac{n}{a} \sum_{1 \leq j \leq \ell} \frac{1}{j}. \tag{14}$$

Therefore, we shall be done if we prove the last inequality in (4). Consider first the case in which $b > m/\sqrt{n}$. Note that, then, we have $\ell = \min\{a, \lceil n/a \rceil\} = \lceil n/a \rceil$, and hence $\ell \geq n/a$. This implies that $\ln b^2 n/m^2 = \ln a^2/n \geq \ln a/\ell$, and hence, by the second bound in (2), we have that (14) is at most

$$\frac{n}{a}(\ln a + \gamma) - \frac{1}{2} \ln \frac{a}{\ell} + \frac{1}{2}.$$

Let us now consider the case in which $b \leq m/\sqrt{n}$. Then $\ell = a$. Notice that, moreover, $\lceil bn/m \rceil = bn/m = a$. Therefore the first bound in (2) becomes

$$\frac{n}{a} + \frac{n}{a} \sum_{2 \leq j \leq a} \frac{1}{j} = \frac{n}{a} \sum_{1 \leq j \leq a} \frac{1}{j} < \frac{n}{a}(\ln(a+1) + \gamma).$$

This completes the proof of the second inequality in (4), and Corollary 2 follows.

3 Lower bounds

We work with families of translations of a-element subsets of $\mathbb{Z}/n\mathbb{Z}$.

If $A = \{\bar{0}, \bar{1}, \dots, \overline{a-1}\} \subset \mathbb{Z}/n\mathbb{Z}$ and $\mathcal{C} = \{A + t \mid t \in \mathbb{Z}/n\mathbb{Z}\}$, where $A + t = \{x + t \mid x \in A\}$, then $\alpha(\mathbb{Z}/n\mathbb{Z}, \mathcal{C}) = \lceil n/a \rceil$.

In the other direction, we have the following proposition.

Proposition 6 *There is $a_0 \in \mathbb{N}$ such that if $n > a \geq a_0$ then, for some $A \subset F := \mathbb{Z}/n\mathbb{Z}$ with $|A| = a$, the family $\mathcal{C} = \{A + t \mid t \in \mathbb{Z}/n\mathbb{Z}\}$ is such that*

$$\alpha(F, \mathcal{C}) > k_0 = \left(1 - \frac{12 \ln \ln a}{\ln a}\right) \frac{\ln a}{\ln \dfrac{1}{1 - a/n}} > \left(1 - \frac{12 \ln \ln a}{\ln a}\right) \left(\frac{n}{a} - 1\right) \ln a. \quad (15)$$

An interesting feature of Proposition 6 is that it claims the existence of uniform, regular hypergraphs with large α. More importantly, the parameters $n > a$ are free and (15) gives good estimates regardless of the relation between them. The reader is invited to compare the bounds in (4) and (15) for the cases in which (*i*) $a \sim \ln n$, (*ii*) $a \sim n/\ln n$, and (*iii*) $a \sim n/2$. In the course of answering a question of Tuza, Alon [1] obtained sharp bounds for case (*ii*), although, strictly speaking, the hypergraphs in [1] are not precisely a-uniform (the hyperedges have average cardinality a). We also observe that, in Proposition 6 above, for simplicity, we restrict ourselves to (a, b)-regular hypergraphs with $a = b$.

3.1 Proof of the lower bound

We now prove Proposition 6. The proof is split into two cases, according to the size of a. We deal with the case in which a is large first; the other case will require an additional idea.

3.1.1 Large a

Here, we suppose that $a \geq n/(\ln n)^3$. We consider all the a-element subsets of $\mathbb{Z}/n\mathbb{Z}$, taken with equal probability. Let us estimate the probability that such a set A has k translations that cover F, where k is a given positive integer, i.e., let us estimate the probability $p(n, a, k)$ that there should exist $t_1, \dots, t_k \in \mathbb{Z}/n\mathbb{Z}$ such that $(A + t_1) \cup \dots \cup (A + t_k) = \mathbb{Z}/n\mathbb{Z}$. Note that $1 - p(n, a, k)$ is the probability that $(A + t_1) \cup \dots \cup (A + t_k) \neq \mathbb{Z}/n\mathbb{Z}$ for any $t_1, \dots, t_k \in \mathbb{Z}/n\mathbb{Z}$, i.e., $(A^c + t_1) \cap \dots \cap (A^c + t_k) \neq \emptyset$ for any $t_1, \dots, t_k \in \mathbb{Z}/n\mathbb{Z}$. Observe that, given $t_1, \dots, t_k \in \mathbb{Z}/n\mathbb{Z}$, we have $(A^c + t_1) \cap \dots \cap (A^c + t_k) \neq \emptyset$ if and only if there is $x \in \mathbb{Z}/n\mathbb{Z}$ such that $\{x - t_1, \dots, x - t_k\} \subset A^c$.

Fix $T = \{t_1, \dots, t_k\} \subset \mathbb{Z}/n\mathbb{Z}$.

Claim 7 *Set $r = \lceil n/k^2 \rceil$. There exist $x_1, \dots, x_r \in \mathbb{Z}/n\mathbb{Z}$ such that the sets $B_i = x_i - T = \{x_i - t_1, \dots, x_i - t_k\}$ $(1 \le i \le r)$ are pairwise disjoint.*

PROOF. To prove our claim, take $x_1 = 0$, and suppose that we have chosen $x_1, \dots, x_s$, with $s < n/k^2$, such that $B_1, \dots, B_s$ are pairwise disjoint. Since $s < n/k^2$, clearly $U = \bigcup_{1 \le i \le s}(x_i - T)$ has fewer than n/k elements. Thus the average cardinality of the intersection of U with $x - T$, for $x \in \mathbb{Z}/n\mathbb{Z}$, is strictly smaller than one, whence there is $x_{s+1} \in \mathbb{Z}/n\mathbb{Z}$ such that $x_{s+1} - T$ is disjoint from U. Our claim thus follows by induction. (See also Section 4.) $\square$

Let us now proceed with the proof of our proposition. For any fixed set $T = \{t_1, \dots, t_k\}$, the probability that $x - T \subset A^c$ for some $x \in \mathbb{Z}/n\mathbb{Z}$ is at least the probability that $x_j - T \subset A^c$ for some $j \in [r] := \{1, \dots, r\}$, where the x_j $(1 \le j \le r)$ are fixed and are as given by our claim. This latter probability is $1 - \tilde{p}(n, a, k, r)$, where $\tilde{p}(n, a, k, r)$ is the probability that $x_j - T \not\subset A^c$ for each $j \in [r]$. To estimate $\tilde{p}(n, a, k, r)$, we consider random subsets $\tilde{A} \subset \mathbb{Z}/n\mathbb{Z}$ constructed as follows: let $y \in \tilde{A}$ with probability a/n, independently for all $y \in \mathbb{Z}/n\mathbb{Z}$. The probability that $\tilde{A}$ has m elements is $\binom{n}{m}(a/n)^m(1 - a/n)^{n-m}$, which is maximal for $m = a$, so the probability that $|\tilde{A}| = a$ is at least $1/(n+1)$. With this probability distribution, the events $B_j = x_j - T \not\subset \tilde{A}^c$ $(j \in [r])$ are independent (because the sets B_j are pairwise disjoint), and the probability of each of these events is $1 - (1 - a/n)^k$. Hence the probability that $B_j \not\subset \tilde{A}^c$ for all $j \in [r]$ is $(1 - (1 - a/n)^k)^r$. So we have

$$\begin{aligned}\tilde{p}(n, a, k, r) &= \mathbb{P}\left(\forall j \in [r] \text{ we have } x_j - T \not\subset \tilde{A}^c \;\middle|\; |\tilde{A}| = a\right) \\ &\le (n+1)\,\mathbb{P}(\forall j \in [r] \text{ we have } x_j - T \not\subset \tilde{A}^c) = (n+1)\left(1 - \left(1 - \frac{a}{n}\right)^k\right)^r,\end{aligned}$$

and hence $p(n, a, k)$ is at most

$$\binom{n}{k}\tilde{p}(n, a, k, r) \le \binom{n}{k}(n+1)\left(1 - \left(1 - \frac{a}{n}\right)^k\right)^r \le n^{k+1}\left(1 - \left(1 - \frac{a}{n}\right)^k\right)^r. \tag{16}$$

Let $k = -\beta \ln a / \ln(1 - a/n) \le \beta(n/a)\ln n$, where $\beta \le 1$. We have that the right-hand side of (16) is $n^{k+1}\left(1 - a^{-\beta}\right)^r$, which is less than

$$n^{1+\beta \ln^4 n} \exp\left(-\frac{a^{-\beta} n}{k^2}\right) < \exp\left(-\frac{a^{2-\beta}}{\beta^2 n \ln^2 n} + \ln n\left(1 + \ln^4 n\right)\right).$$

One may check that if $\beta \le (1 - 12(\ln\ln a)/\ln a)$, then $a^{2-\beta}/\beta^2 n \ln^2 n \ge \ln^6 n$ for large n.

Since $\ln^6 n \gg \ln n(1+\ln^4 n)$, we have $p(n,a,k) \ll 1$. Therefore there is $A \subset \mathbb{Z}/n\mathbb{Z}$ with $|A| = a$ such that $(A+t_1) \cup \cdots \cup (A+t_k) \neq \mathbb{Z}/n\mathbb{Z}$ for any $t_1, \ldots, t_k \in \mathbb{Z}/n\mathbb{Z}$. This implies that, for some $A \subset \mathbb{Z}/n\mathbb{Z}$ with $|A| = a$, we have $\alpha(F, \mathcal{C}) > k_0$, where k_0 is as in (15). This completes the proof of Proposition 6 in the case in which $a \geq n/(\ln n)^3$.

3.1.2 Small a

We now deal with the case in which a is 'small', that is, $a < n/(\ln n)^3$. Let

$$b = \lceil a(\ln a + 1) \rceil, \quad r = \lceil \ln a \rceil, \quad \varepsilon = \frac{1}{r}, \quad \delta = \frac{10 \ln \ln a}{\ln a}, \tag{17}$$

and

$$k = \left\lceil \frac{1-\delta}{1+\varepsilon} r (\ln a)^2 \right\rceil. \tag{18}$$

We shall make use of the following claim, to be proved later (see §3.1.2.1).

Claim 8 *There exist $0 \leq y_i < b$ $(1 \leq i \leq a)$ so that if we let $A_0 = \{\overline{y}_1, \ldots, \overline{y}_a\} \subset \{\overline{0}, \overline{1}, \ldots, \overline{b-1}\} \subset \mathbb{Z}/rb\mathbb{Z}$, then*

$$(A_0 + t_1) \cup \cdots \cup (A_0 + t_k) \neq \mathbb{Z}/rb\mathbb{Z}$$

for all $t_1, \ldots, t_k \in \mathbb{Z}/rb\mathbb{Z}$.

We now prove Proposition 6 for $a < n/(\ln n)^3$ assuming Claim 8. Let $\ell = \lfloor n/(r+1)b \rfloor$, and let y_i $(1 \leq i \leq a)$ be as in Claim 8 above. Put $A_0 = \{\overline{y}_1, \ldots, \overline{y}_a\} \subset \mathbb{Z}/rb\mathbb{Z}$, and let $A = \{y_1 \bmod n, \ldots, y_a \bmod n\} \subset \mathbb{Z}/n\mathbb{Z}$. We claim that

$$(A+s_1) \cup \cdots \cup (A+s_m) = \mathbb{Z}/n\mathbb{Z} \tag{19}$$

implies that $m > \ell(k+1)$. To prove this claim, suppose (19) holds for some $s_1, \ldots, s_m \in \mathbb{Z}/n\mathbb{Z}$. For $0 \leq j < \ell$, let

$$B_j = \{jb(r+1) + q \mid 0 \leq q < rb\} \text{ and } I_j = \{i \leq m \mid (A+s_i) \cap B_j \neq \emptyset\}.$$

The sets I_j are pairwise disjoint, since the diameter of $A+s_i$ is at most b, and the distance between B_j and B_{j+1} is $b+1$. Moreover, each I_j must have at least $k+1$ elements, since

$$\bigcup_{i \in I_j} (A+s_i) \supset B_j$$

implies that

$$\bigcup_{i \in I_j} (A_0 + \overline{s}_i) = \mathbb{Z}/rb\mathbb{Z},$$

which, by the choice of $A_0 = \{\overline{y}_1, \ldots, \overline{y}_a\}$, implies that $|I_j| > k$.

To finish the proof we just notice that

$$\begin{aligned}\ell(k+1) &= \left\lfloor \frac{n}{(r+1)b} \right\rfloor \left(\left\lceil \frac{1-\delta}{1+\varepsilon} r (\ln a)^2 \right\rceil + 1 \right) \\ &> \left(\frac{n}{(r+1)b} - 1 \right) \frac{1-\delta}{1+\varepsilon} (\ln a)^3 > \frac{n}{b(r+1)} (\ln a)^3 \left(1 - \frac{11 \ln\ln a}{\ln a} \right) \\ &\qquad > \frac{n}{a} (\ln a) \left(1 - \frac{12 \ln\ln a}{\ln a} \right) \geq \left(1 - \frac{12 \ln\ln a}{\ln a} \right) \frac{\ln a}{\ln \dfrac{1}{1 - a/n}},\end{aligned}$$

for large enough a. $\square$

3.1.2.1 Proof of Claim 8. In order to prove Claim 8, we consider a random subset A_0 of $\{\overline{0}, \overline{1}, \ldots, \overline{b-1}\}$, with each element present in A_0 independently with probability a/b. The probability that A_0 has m elements is $\binom{b}{m}(a/b)^m(1-a/b)^{b-m}$, which is maximal for $m = a$. Therefore, the probability that such a set A_0 has a elements is at least $1/(b+1)$. As before, we shall condition on the event $|A_0| = a$ later in the proof.

Let us fix $t_1, \ldots, t_k$ and let us estimate from above the probability that

$$(A_0 + t_1) \cup \cdots \cup (A_0 + t_k) = \mathbb{Z}/rb\mathbb{Z}. \tag{20}$$

Put $T = \{t_1, \ldots, t_k\}$, and observe that (20) occurs if and only if for all $x \in \mathbb{Z}/rb\mathbb{Z}$, the set $x - T = \{x - t_1, \ldots, x - t_k\}$ meets A_0. Let

$$s = \left\lceil \frac{\varepsilon}{1+\varepsilon} rb \right\rceil \quad \text{and} \quad s_0 = \left\lceil \frac{s}{k^2} \right\rceil. \tag{21}$$

We now prove the following two facts (cf. Claim 7):

(*) There are $x_1, \ldots, x_s$ such that, for each i, we have

$$|(x_i - T) \cap \{\overline{0}, \overline{1}, \ldots, \overline{b-1}\}| \leq (1+\varepsilon)\frac{k}{r}.$$

.

(**) There are $\tilde{x}_1, \ldots, \tilde{x}_{s_0} \in \{x_1, \ldots, x_s\}$ such that the sets $\tilde{x}_j - T$ $(1 \leq j \leq s_0)$ are pairwise disjoint.

To prove (*), it suffices to observe that the average number of elements in $(x - T) \cap \{\overline{0}, \overline{1}, \ldots, \overline{b-1}\}$ $(x \in \mathbb{Z}/rb\mathbb{Z})$ is k/r. The proof of (**) is similar to the proof of Claim 7: suppose we have $\tilde{x}_1, \ldots, \tilde{x}_{s_1} \in \{x_1, \ldots, x_s\}$ such that the sets $\tilde{x}_j - T$ $(1 \leq j \leq s_1)$ are pairwise disjoint, but $s_1 < s/k^2$. Then

$$\left| \bigcup_{1 \leq j \leq s_1} (\tilde{x}_j - T) \right| = s_1 k < \frac{s}{k}. \tag{22}$$

If we select $x_i \in \{x_1, \ldots, x_s\}$ uniformly at random, then the probability that a fixed element z in $\mathbb{Z}/rb\mathbb{Z}$ belongs to $x_i - T$ is at most k/s, because $k = |T|$ translates of T contain z. Because of (22), the expected cardinality of

$$(x_i - T) \cap \bigcup_{1 \le j \le s_1} (\tilde{x}_j - T)$$

is strictly smaller than 1. Therefore the sequence $\tilde{x}_1, \ldots, \tilde{x}_{s_1}$ may be extended with a new element $x_i \in \{x_1, \ldots, x_s\}$. This completes the proof of (**).

For the remainder of the proof, we concentrate our attention on the $\tilde{x}_j$ in (**). One may easily check that the probability that $(\tilde{x}_j - T) \cap A_0 \neq \emptyset$ occurs for all $1 \le j \le s_0$ is at most

$$\left(1 - \left(1 - \frac{a}{b}\right)^{(1+\varepsilon)k/r}\right)^{s_0}.$$

Therefore the probability that, for some $T = \{t_1, \ldots, t_k\} \subset \mathbb{Z}/rb\mathbb{Z}$, we have $(\tilde{x}_j - T) \cap A_0 \neq \emptyset$ for all $1 \le j \le s_0$, conditioned on the event $|A_0| = a$, is at most

$$(b+1)(rb)^k \left(1 - \left(1 - \frac{a}{b}\right)^{(1+\varepsilon)k/r}\right)^{s_0}. \tag{23}$$

We now estimate (23) in parts. In the calculations below, we tacitly assume that a is larger than a suitable constant. Since $1 - a/b \ge 1 - 1/(\ln a + 1)$, we have

$$\left(1 - \frac{a}{b}\right)^{(1+\varepsilon)k/r} \ge \exp\left(-(1+\varepsilon)\frac{k}{r(\ln a)}\right). \tag{24}$$

We have

$$(1+\varepsilon)\frac{k}{r(\ln a)} \ge (1-\delta)\ln a. \tag{25}$$

Putting together (24) and (25), we have

$$\left(1 - \left(1 - \frac{a}{b}\right)^{(1+\varepsilon)k/r}\right)^{s_0} \le \left(1 - a^{-1+\delta}\right)^{s_0} \le \exp\left(-a^{-1+\delta}\frac{\varepsilon rb}{(1+\varepsilon)k^2}\right). \tag{26}$$

Very generously, we have $rb/k^2 \ge a/3(\ln a)^4$. Therefore, again generously, we have

$$a^{-1+\delta}\frac{\varepsilon rb}{(1+\varepsilon)k^2} \ge \frac{a^\delta}{7(\ln a)^5}. \tag{27}$$

On the other hand, a crude estimate gives

$$(b+1)(rb)^k \le \exp\left(4(\ln a)^4\right). \tag{28}$$

Putting together (26), (27), and (28), we see that the quantity in (23) is bounded from above by

$$\exp\left(-\frac{a^\delta}{7(\ln a)^5} + 4(\ln a)^4\right). \tag{29}$$

Because of our choice of δ (see (17)), we have that $a^{\delta} = (\ln a)^{10} \gg (\ln a)^{9}$ as $a \to \infty$, and hence the quantity in (29) is < 1 for any large enough a. We conclude that the probability that A_0 will do in Claim 8 is positive, and hence the claim is proved. $\square$

3.2 Constructive lower bounds

Recall that we prove the existence of systems $\mathcal{C}$ with large $\alpha(F, \mathcal{C})$ by taking $F = \mathbb{Z}/n\mathbb{Z}$ and considering translates $A + t$ ($t \in \mathbb{Z}/n\mathbb{Z}$) for suitable random sets $A \subset \mathbb{Z}$. As already observed by Alon [1], if we take n to be a prime power q and let $A \subset F = \mathrm{GF}(q)$ be the set of squares in $\mathrm{GF}(q)$, then

$$\mathcal{C} = \{A + t \mid t \in \mathrm{GF}(q)\}$$

is an (a, a)-regular system for $a = (q-1)/2$ and

$$\alpha(F, \mathcal{C}) \geq \left(\frac{1}{2} - o(1)\right) \lg q, \tag{30}$$

where we write lg for the logarithm to the base 2. The bound in (30) follows from the following result, which we quote from [2] (see Lemma 9, Chapter 13) without proof. Let χ be the quadratic character on $\mathrm{GF}(q)$, so that $\chi(x) = x^{(q-1)/2}$ ($x \in \mathrm{GF}(q)$). We have $\chi(x) \in \{\pm 1, 0\}$, with $\chi(x) = 0$ if $x = 0$ and $\chi(x) = 1$ if and only if x is a square in $\mathrm{GF}(q) \setminus \{0\}$.

Lemma 9 *If $T \subset \mathrm{GF}(q)$ and $k = |T|$, then*

$$\left| q - \sum_{x \notin T} \prod_{t \in T} (1 - \chi(x - t)) \right| \leq \Big((k-2)2^{k-1} + 1\Big) q^{1/2} + k2^{k-1}. \tag{31}$$

Lemma 9 is in fact a consequence of a well known estimate of Weil for character sums (see [2]). To deduce (30) from Lemma 9, let $T \subset \mathrm{GF}(q)$ be an arbitrary set with $k = |T| = \lfloor (1/2) \lg q - \lg \lg q \rfloor$. The element $x \in \mathrm{GF}(q)$ will not be covered by the translates $A + t$ ($t \in T$) if and only if $x - T$ fails to meet A, that is, $x - t$ is not a square for any $t \in T$. Now, the number of such x is

$$2^{-k} \sum_{x \notin T} \prod_{t \in T} (1 - \chi(x - t)) \,.$$

Since by the choice of k we have

$$2^{-k} q > \frac{1}{2}(k-2) q^{1/2} + 2^{-k} q^{1/2} + \frac{k}{2},$$

the existence of such an x follows from (31).

Finally, let us observe that our lower bound k_0 in Proposition 6 for the case in which $a = n/2$ (suppose n even for simplicity) is

$$k_0 = \left(1 - \frac{12 \ln \ln n}{\ln n}\right) \lg n.$$

Furthermore, the upper bound for $\alpha(F, \mathcal{C})$ in Corollary 2 for this case is $\lg n + 1$. Therefore Alon's construction is off only by a factor of 2.

4 Packings

We briefly consider the problem of finding large packings in regular families. Suppose $\mathcal{C}$ is a family of subsets of a set F. Let $\beta(F, \mathcal{C})$ be the maximal integer r for which there exist pairwise disjoint sets $C_1, \ldots, C_r \in \mathcal{C}$.

We prove the following proposition.

Proposition 10 *Suppose $\mathcal{C}$ is an (a, b)-regular family on an n-element set F. Then*

$$\frac{n}{a^2} \leq \beta(F, \mathcal{C}) \leq \frac{n}{a}. \tag{32}$$

Proposition 10 follows from the following lemma.

Lemma 11 *Let $\mathcal{C}$ be an (a, b)-regular family of sets on a set F. Given a subset $\widetilde{\mathcal{C}} \subset \mathcal{C}$ with r elements, it is possible to find a subset $\mathcal{B} \subset \widetilde{\mathcal{C}}$ of disjoint sets with at least r/ab elements.*

To prove Proposition 10, observe that if $\mathcal{C}$ is as in the statement of that result, then we may take $\widetilde{\mathcal{C}} = \mathcal{C}$. Note that then $r = |\widetilde{\mathcal{C}}| = |\mathcal{C}| = bn/a$, and hence $r/ab = n/a^2$, and the lower bound in (32) follows. The upper bound in (32) is obvious.

We now prove Lemma 11.

PROOF. (Proof of Lemma 11) Let s be the maximal number of pairwise disjoint members in $\widetilde{\mathcal{C}}$. Suppose for a contradiction that $s < r/ab$, and let $C_1, \ldots, C_s \in \widetilde{\mathcal{C}}$ be such a maximal collection. Let $A = \bigcup_{1 \leq j \leq s} C_j$. We have $|A| = as$, so the number of members of $\widetilde{\mathcal{C}}$ that intersect A is at most $abs < r = |\widetilde{\mathcal{C}}|$. Therefore there is $C_{s+1} \in \widetilde{\mathcal{C}}$ that is disjoint from all the C_j $(1 \leq j \leq s)$, which contradicts the maximality of $C_1, \ldots, C_s$. $\square$

4.1 An example

We now observe that the bounds in Proposition 10 cannot be substantially improved without further hypotheses. Indeed, given a prime power q we may take F to be the projective plane over the finite field $\mathrm{GF}(q)$, and $\mathcal{C}$ to be the collection of lines of F. Then $|F| = q^2 + q + 1$, the system $\mathcal{C}$ is $(q+1, q+1)$-regular, and $\beta(F, \mathcal{C}) = 1$. Notice that the lower bound in (32) tells us that $\beta(F, \mathcal{C}) \geq (q^2 + q + 1)/(q+1)^2 \to 1$ as $q \to \infty$.

More generally, given a positive integer r we may take $F_r = F \times \{1, 2, \dots, r\}$ and $\mathcal{C}_r = \{L \times \{j\} \mid L \in \mathcal{C},\ 1 \leq j \leq r\}$ (i.e., F_r is the union of r disjoint copies of the projective plane over $\mathrm{GF}(q)$ and $\mathcal{C}_r$ is the collection of lines in these copies). We have $|F_r| = r(q^2+q+1)$, the system $\mathcal{C}_r$ is $(q+1, q+1)$-regular, and $\beta(F, \mathcal{C}) = r$, which is close to the lower bound $r(q^2+q+1)/(q+1)^2$ given by (32) provided q is large. We can use these examples in order to show that, given sequences of integers (a_k) and (n_k), with $a_k \to \infty$ and $n_k/a_k^2 \to \infty$ as $k \to \infty$, there exist sequences $(\tilde{a}_k)$ and $(\tilde{n}_k)$ such that $\tilde{a}_k/a_k$ and $\tilde{n}_k/n_k$ tend to 1 as $k \to \infty$ and for which there exist $F^{(k)}$ and $\mathcal{C}^{(k)}$ such that $\mathcal{C}^{(k)}$ is an $(\tilde{a}_k, \tilde{a}_k)$-regular family of sets on $F^{(k)}$, where $|F^{(k)}| = \tilde{n}_k$, and

$$\lim_{k\to\infty} \beta(F^{(k)}, \mathcal{C}^{(k)}) \left(\frac{\tilde{n}_k}{\tilde{a}_k^2}\right)^{-1} = 1$$

(here we use the fact that there is always a prime between x and $(1 + o(1))x$, which follows from the prime number theorem).

5 Concluding remarks

If a is a positive integer, let

$$\alpha(a, n) = \max \alpha(F, \mathcal{C}),$$

where the maximum is taken over all (a, b)-regular families of sets $\mathcal{C}$ on an n-element set F, and $b \geq 1$ is arbitrary. Put

$$f(a) = \limsup_{n\to\infty} \frac{a}{n} \alpha(a, n).$$

Our results imply that, for any large enough fixed a, we have

$$\ln a - 12 \ln\ln a \leq f(a) \leq \sum_{1 \leq k \leq a} \frac{1}{k} = \ln a + \gamma + O\left(\frac{1}{a}\right). \tag{33}$$

Notice that the upper bound for $f(a)$ in (33) above holds for every a, by Corollary 2.

Consider the case in which $a = 2$, that is, the case of regular graphs. It is not difficult to show that $(2/n)\alpha(2, n) = 4/3 + o(1)$ as $n \to \infty$, so that $f(2) = 4/3$. Indeed, for the lower bound, just take for $\mathcal{C}$ a collection of, say, k vertex disjoint triangles on an $3k$-element set F. Then $\alpha(F, \mathcal{C}) = 2k$ and we conclude that $f(2) \geq 4/3$.

To prove the upper bound, we show that any b-regular graph G $(b > 0)$ must contain a matching that covers at least $2/3$ of its vertices. Let M be a maximum matching in G, and suppose U is the set of vertices that are covered by M. Suppose for a contradiction that $|U| < (2/3)n$, where $n = |V(G)|$. Let $W = U^c = V(G) \setminus U$. Let the number of neighbours in W of a vertex u in U be the *W-degree $d_W(u)$ of u*. The average W-degree of a vertex in U is $|W|b/|U|$. Thus there is an edge $e \in M$ whose endpoints x and y are such that

$$d_W(x) + d_W(y) \geq 2|W|b/|U| > b. \tag{34}$$

Note that $d_W(x)$, $d_W(y) < b$ (because of the edge $e = \{x, y\} \subset U$). Therefore (34) implies that $d_W(x)$, $d_W(y) \geq 2$. But then there exist distinct vertices x', $y' \in W$, with x' adjacent to x and y' adjacent to y'. Now observe that $M \setminus \{e\} \cup \{xx', yy'\}$ is a larger matching than M, which contradicts the maximality of $|M|$. This contradiction shows that M does indeed cover $(2/3)n$ vertices of G, and hence $\alpha(F, \mathcal{C}) \leq 2n/3$. This implies $f(2) \leq 4/3$.

Problem 12 *Determine $f(a)$ for all $a \geq 3$.*

Finally, we believe that it would be very interesting to improve on the constructive lower bounds (see Section 3.2).

References

[1] N. Alon: Transversal numbers of uniform hypergraphs, *Graphs Combin.* **6** (1990), 1–4.

[2] B. Bollobás: Random Graphs, *Academic Press*, London, 1985.

[3] G. Cohen, I. Honkala, S. Litsyn and A. Lobstein: Covering Codes, *North-Holland*, Amsterdam, 1997.

[4] Z. Füredi: Matchings and covers in hypergraphs, *Graphs Combin.* **4** (1988), 115–206.

[5] Z. Füredi, G.J. Székely and Z. Zubor: On the lottery problem, *J. Combin. Des.* **4** (1996), 5–10.

[6] H. Hämäläinen, I. Honkala, S. Litsyn and P. Östergård: Football pools — a game for mathematicians, *Amer. Math. Monthly* **102** (1995), 579–588.

[7] D.S. Johnson: Approximation algorithms for combinatorial problems, *J. Comput. System Sciences* **9** (1974), 256–298.

[8] L. Lovász: On the ratio of optimal integral and fractional covers, *Discrete Math.* **13** (1975), 383–390.

[9] V. Rödl: On a packing and covering problem, *Europ. J. Combin.* **6** (1985), 69–78.

[10] S.K. Stein: Two combinatorial covering problems, *J. Comb. Th.*, Ser. A **16** (1974), 391–397.