COUNTING SETS WITH SMALL SUMSET, AND THE CLIQUE NUMBER OF RANDOM CAYLEY GRAPHS

BEN GREEN[1]

Given a set $A \subseteq \mathbb{Z}/N\mathbb{Z}$ we may form a Cayley sum graph G_A on vertex set $\mathbb{Z}/N\mathbb{Z}$ by joining i to j if and only if $i+j \in A$. We investigate the extent to which performing this construction with a random set A simulates the generation of a random graph, proving that the clique number of G_A is almost surely $O(\log N)$. This shows that Cayley sum graphs can furnish good examples of Ramsey graphs. To prove this result we must study the specific structure of set addition on $\mathbb{Z}/N\mathbb{Z}$. Indeed, we also show that the clique number of a random Cayley sum graph on $\Gamma = (\mathbb{Z}/2\mathbb{Z})^n$ is almost surely *not* $O(\log|\Gamma|)$.

1. Introduction

There are many papers in which a graph is constructed from a set in something like the following manner.

Construction 1. Let Γ be a finite abelian group with cardinality N and let $A \subseteq \Gamma$. Define G_A to be the graph on vertex set Γ in which $ij \in E(G_A)$ precisely when $i+j \in A$.

For any positive integer m write $\mathbb{Z}_m = \mathbb{Z}/m\mathbb{Z}$. In this paper we will be interested in the two particular groups $\Gamma = \mathbb{Z}_N$ and $\Gamma = \mathbb{Z}_2^n$.

Example. Suppose that N is prime and let A be the set of quadratic residues modulo N. The graph G_A is then called the *Paley sum graph* of order N, and is denoted by P_N. It is generally thought that the Paley graphs

---

behave much like random graphs from the binomial model $G(N, 1/2)$, though in many respects this statement is far from being proven. We remind the reader that in this model all (labelled) graphs on N vertices occur with equal probability. Equivalently, a graph from $G(N, 1/2)$ may be generated by picking each possible edge ij independently at random with probability $1/2$. See [6] for the basic definitions of random graph theory (and much more).

In the light of the above example it seems natural to address the following question.

Question 2. To what extent does choosing a random subset $A \subseteq \Gamma$ and forming G_A simulate the selection of a random graph on N vertices?

We call a graph generated using this model a *random Cayley sum graph*.

The reader may have already noticed that there is one sense in which a random Cayley sum graph is definitely *not* like a random graph – it is always regular. However in a recent paper of Krivelevich, Sudakov, Vu and Wormald [13] a good quantity of evidence was collected to support the claim that a random regular graph of degree $N/2$ (say) behaves in many ways almost exactly like a random graph from the binomial model $G(N, 1/2)$.

In this paper we will investigate just one aspect of Question 2, the clique number of a random Cayley sum graph. There are clearly other ways in which we might compare a random Cayley sum graph to a random graph. An investigation of the second largest eigenvalue, for example, leads naturally to some classical questions on random trigonometric polynomials. We discuss this matter in a separate paper [10].

This is really a paper in combinatorial number theory. To see why, let $A \subseteq \Gamma$ and consider what it means for $X \subseteq \Gamma$ to span a clique in G_A. We must have $i + j \in A$ whenever i, j are distinct elements of X. Now the set $\{i + j : i, j \in X, i \neq j\}$ is called the restricted sumset of X with itself and is normally written $X \hat{+} X$. Thus we see that X spans a clique in G_A if and only if $X \hat{+} X \subseteq A$.

Now if X is highly structured, for example if it is an arithmetic progression, then $|X \hat{+} X|$ is rather small compared to $|X|$ and the probability that G_A contains the clique spanned by X will be quite large. This does not mean, however, that the clique number of a random Cayley graph is very large since there are rather few arithmetic progressions.

Let $k, m \in \mathbb{N}$, and make the following definition.

Definition 3. $S_k^m(\Gamma)$ is the collection of sets $X \subseteq \Gamma$ with $|X| = k$ and $|X \hat{+} X| = m$.

Observe that $S_k^m(\Gamma)$ is empty if $m < k-1$. Now one can easily write down a formula for $E(\Gamma, k)$, the expected number of k-cliques in a random Cayley graph on Γ. Indeed one has

$$
\begin{aligned}
E(\Gamma; k) &= \sum_{X \subseteq \Gamma : |X|=k} \mathbb{P}(X \text{ is a clique in } G_A) \\
&= \sum_{m \geq k-1} \sum_{X \in S_k^m(\Gamma)} \mathbb{P}(X \hat{+} X \subseteq A) \\
&= \sum_{m \geq k-1} \sum_{X \in S_k^m(\Gamma)} 2^{-m} \\
&= \sum_{m \geq k-1} |S_k^m(\Gamma)| 2^{-m}.
\end{aligned}
$$

Now suppose we could show, for some function $F = F_\Gamma : \mathbb{N} \to \mathbb{N}$, that $E(\Gamma, F(N)) = o(1)$. Writing ω for the clique number and Ω_k for the number of cliques of size k we would then have

$$
\mathbb{P}(\omega \geq F(N)) = \mathbb{P}(\Omega_{F(N)} \geq 1) \leq E(\Gamma, F(N)) = o(1)
$$

by Markov's inequality. That is, the clique number of a random Cayley graph on Γ is almost surely at most $F(N)$. As a result of these observations our attention is drawn to the problem of finding effective upper bounds for $|S_k^m(\Gamma)|$.

Further notation and background. Let $|\Gamma| = N$ and consider the sample space Ω_N consisting of all graphs on N vertices. If $E = E_N$ is an event in this sample space then we write

- $\mathbb{P}_{\text{ran}}^{(N)}(E)$ for the probability that a random graph from $G(N, 1/2)$ satisfies E;
- $\mathbb{P}_{\text{reg}}^{(N)}(E)$ for the probability that a graph selected from $\mathcal{G}_{N, \lfloor N/2 \rfloor}$ satisfies E, where $\mathcal{G}_{N, \lfloor N/2 \rfloor}$ puts the uniform distribution on all the $\lfloor N/2 \rfloor$-regular graphs on N labelled vertices;
- $\mathbb{P}_\Gamma(E)$ for the probability that if $A \subseteq \Gamma$ is chosen randomly, then the Cayley sum graph G_A satisfies E.

The following well-known result concerning the clique number of a random graph (see, for example, [6]) sets our later results in context. In this paper, all logs are to the base 2 unless otherwise specified.

Theorem 4. *For any fixed $\epsilon > 0$ we have*

$$
\lim_{N \to \infty} \mathbb{P}_{\text{ran}}^{(N)} \left((2 - \epsilon) \log N \leq \omega \leq (2 + \epsilon) \log N \right) = 1.
$$

In [13] the same result is proved for random regular graphs:

Theorem 5. *For any fixed $\epsilon > 0$ we have*

$$\lim_{N \to \infty} \mathbb{P}_{\mathrm{reg}}^{(N)} \left((2 - \epsilon) \log N \le \omega \le (2 + \epsilon) \log N \right) = 1.$$

Questions equivalent to estimating the clique number of a random Cayley sum graph appear in various places in the literature[2], such as [1–3, 7, 15]. To the best of my knowledge the following result, which is essentially given in [1], is the best known.

Theorem 6. *There is an absolute constant C such that*

$$\lim_{N \to \infty} \mathbb{P}_{\mathbb{Z}_N} \left(\omega \le C (\log N)^2 \right) = 1.$$

A similar result is mentioned in [3], where it is remarked that it would be interesting to try and improve on it, and also, without proof, in [15].

Let us conclude this section by making a remark concerning our terminology. We use the term *Cayley sum graph* to make it clear that the objects we are considering are very slightly different from what are normally called Cayley graphs. In an abelian group Γ a (true) Cayley graph is formed as follows. Take a symmetric set A, that is to say a set in which $a \in A$ implies that $-a \in A$. Form a graph on vertex set Γ by joining i to j if and only if $i - j \in A$. The group Γ acts on the resulting graph in an obvious way, and for this reason true Cayley graphs are perhaps more natural than Cayley sum graphs. From the rather number-theoretical point of view of this paper, however, the treatment of Cayley sum graphs seems to be a little tidier. The methods of this paper apply to true Cayley graphs with minimal modifications. When $\Gamma = \mathbb{Z}_2^n$, of course, the two types of graph actually coincide.

2. Statement of results

The main result of this paper is the following improvement of Theorem 6.

Theorem 7. *We have*

$$\lim_{N \to \infty} \mathbb{P}_{\mathbb{Z}_N} \left(\omega \le 160 \log N \right) = 1.$$

An immediate consequence of Theorem 7 is the following result in Ramsey theory, for which we do not know an independent proof:

---

[2] The author would like to thank Benny Sudakov for drawing his attention to the two papers [1, 3].

Theorem 8. *For all sufficiently large integers N there exists a set $A \subseteq \mathbb{Z}_N$ for which the Cayley sum graph G_A has no cliques or independent sets of size $160 \log N$.*

The behaviour of random Cayley sum graphs on $\mathbb{Z}_2^n$ is different, reflecting the fact that this group has much richer subgroup structure. We will prove the following.

Theorem 9. *There are absolute constants $C_1, C_2 > 0$ such that*

$$\lim_{N \to \infty} \mathbb{P}_{\mathbb{Z}_2^n} \left(C_1 \log N \log \log N \leq \omega \leq C_2 \log N \log \log N \right) = 1,$$

where $N = 2^n = |\mathbb{Z}_2^n|$.

We make no effort to optimise the values of C_1 and C_2 in this result.

3. Freiman isomorphism classes

A vital notion in this paper is that of a *Freiman isomorphism*. Let s be a positive integer, let A and B be subsets of (possibly different) abelian groups and let $\phi : A \to B$ be a map. Then we say that ϕ is a Freiman s-homomorphism if whenever $a_1, \ldots, a_s, a_1', \ldots, a_s' \in A$ satisfy

$$a_1 + a_2 + \cdots + a_s = a_1' + a_2' + \cdots + a_s' \tag{1}$$

we have

$$\phi(a_1) + \phi(a_2) + \cdots + \phi(a_s) = \phi(a_1') + \phi(a_2') + \cdots + \phi(a_s'). \tag{2}$$

If ϕ has an inverse which is also a s-homomorphism then we say that it is a Freiman s-isomorphism, and write $A \cong_s B$. This is the case precisely if (1) holds for any a_i, a_i' satisfying (2).

This is a standard and very convenient terminology in additive number theory. One must, however, be slightly careful: we invite the reader to consider the obvious map $\phi : \{0, 1\} \to \mathbb{Z}_2$, which is a one-to-one Freiman 2-homomorphism but not a Freiman isomorphism.

In this paper we will mostly, but not exclusively, be concerned with Freiman 2-homomorphisms. From now on we will refer to these simply as homomorphisms.

In order to discuss the theory of Freiman homomorphisms as cleanly as possible we introduce some further definitions. These are very much in the

spirit of [12], and indeed Lemma 11 below is a crude (though slightly more general) version of one of the results of that paper.

Let F be a field, let k be a positive integer and let W be a vector space over F. Let V_k be a k-dimensional vector space over F with basis $\{e_1,\ldots,e_k\}$. For any sequence $\sigma=(a_1,\ldots,a_k)$ of k elements of W we can define a linear map $\phi_\sigma : V_k \to W$ by

$$\phi_\sigma\left(\sum \lambda_i e_i\right) = \sum \lambda_i a_i.$$

If $v \in \ker\phi_\sigma$ then we say that v is *satisfied by* σ. Let $s \geq 2$ be an integer. Call any vector of the form

$$e_{i_1} + \cdots + e_{i_s} - e_{j_1} - \cdots - e_{j_s},$$

where the is and js need not be distinct, an *s-relation*. Write $\mathcal{R}_s$ for the set of all s-relations in V_k. If $S \subseteq V_k$ write $\mathrm{Span}(S)$ for the F-linear span of the elements of S.

Now for any subset S and any subspace U of a vector space one has $S \cap \mathrm{Span}(S \cap U) = S \cap U$. Thus in particular

$$(3) \qquad \mathcal{R}_s \cap \mathrm{Span}\left(\mathcal{R}_s \cap \ker\phi_\sigma\right) = \mathcal{R}_s \cap \ker\phi_\sigma.$$

That is to say, the set $\mathcal{R}_s \cap \ker\phi_\sigma$ of s-relations satisfied by σ can be completely recovered from the subspace spanned by it.

Lemma 10. *Suppose that* $\sigma = (a_i)_{i=1}^k$ *and* $\sigma' = (a_i')_{i=1}^k$ *are two sequences of distinct elements of W, and suppose that* $\mathrm{Span}\left(\mathcal{R}_s \cap \ker\phi_\sigma\right) = \mathrm{Span}\left(\mathcal{R}_s \cap \ker\phi_{\sigma'}\right)$. *Then the sets* $A = \{a_1,\ldots,a_k\}$ *and* $A' = \{a_1,\ldots,a_k\}$ *are s-isomorphic.*

Proof. By (3) we have $\mathcal{R}_s \cap \ker\phi_\sigma = \mathcal{R}_s \cap \ker\phi_{\sigma'}$. That is, the set of s-relations satisfied by σ and σ' are the same. An s-isomorphism between A and A' is therefore given by the map $a_i \mapsto a_i'$. ∎

Lemma 11. *The number of s-isomorphism classes of subsets of W of size k is at most k^{2sk}.*

Proof. Given any set $A = \{a_1,\ldots,a_k\} \subseteq W$ we may form a sequence $\sigma = (a_1,\ldots,a_k)$. By Lemma 10 the number of s-isomorphism classes of sets A is at most the number of subspaces of V_k spanned by s-relations. There are at most k^{2s} of these s-relations, and any such subspace is spanned by some subset of at most k of them. The required bound follows. ∎

Now if W, W' are two vector spaces over F and if $A \subseteq W$ then the space $\mathrm{Hom}_s(A, W')$ of Freiman s-homomorphisms $\phi : A \to W'$ is a vector space

over F. We call $r_F(A) = \dim \operatorname{Hom}_2(A, F) - 1$ the *Freiman dimension* of A. It is easy to see that $\dim \operatorname{Hom}_2(A, W') = (r_F(A) + 1)^{\dim W'}$. In a short while we will prove the main result of this section, Lemma 13. Let us first prepare the ground with the following lemma from linear algebra.

Lemma 12. *Let F be a field, let A be a finite non-empty set, and let Φ be a subspace of Ψ, the vector space of all maps from A to F. Write $d = \dim \Phi$. Then there exists a subset $\{a_1, \ldots, a_d\} \subseteq A$ with the following properties.*
(i) For any $f_1, \ldots, f_d \in F$ there is a unique $\phi \in \Phi$ with $\phi(a_i) = f_i$ $(i = 1, \ldots, d)$.
(ii) If $a \in A$ then there are $\lambda_1, \ldots, \lambda_d \in F$ (depending only on a) such that $\phi(a) = \lambda_1 \phi(a_1) + \cdots + \lambda_d \phi(a_d)$ for any $\phi \in \Phi$. If the constant map, which sends everything in A to 1_F, lies in Φ then we may assume that $\lambda_1 + \cdots + \lambda_d = 1$.

Proof. Let $A = \{a_1, \ldots, a_k\}$. Choose as a basis for Ψ the set $\{\alpha_1, \ldots, \alpha_k\}$ of maps defined by the conditions $\alpha_i(a_j) = \delta_{ij}$. Now we may find some $k - d$ of the α_i which, when taken together with Φ, span the whole of Ψ. Relabeling if necessary we may assume that these elements are $\alpha_{d+1}, \ldots, \alpha_k$, so that

$$(4) \qquad\qquad \Psi = \Phi \oplus \operatorname{Span}(\alpha_{d+1}, \ldots, \alpha_k).$$

It follows that for any $f_1, \ldots, f_d \in F$ we may find $\phi \in \Phi$ and $\mu_{d+1}, \ldots, \mu_k \in F$ so that

$$f_1 \alpha_1 + \cdots + f_d \alpha_d = \phi + \mu_{d+1} \alpha_{d+1} + \cdots + \mu_k \alpha_k,$$

implying that $\phi(a_i) = f_i$ $(i = 1, \ldots, d)$. To see that ϕ is unique, suppose that $\phi'(a_i) = \phi''(a_i)$ for some $\phi', \phi'' \in \Phi$ and all $i = 1, \ldots, d$. Then $\phi' - \phi'' \in \operatorname{Span}(\alpha_{d+1}, \ldots, \alpha_k)$ and thus $\phi' - \phi'' = 0$ since the sum in (4) is direct.

As for assertion (ii), observe that any $x \in A$ induces a linear functional $t_x : \Phi \to F$ via the rule $t_x(\phi) = \phi(x)$. For a fixed $a \in A$ the $d + 1$ linear functionals $t_a, t_{a_1}, \ldots, t_{a_d}$ are linearly dependent and therefore we can find $\mu, \mu_1, \ldots, \mu_d \in F$, not all zero, so that

$$(5) \qquad\qquad \mu \phi(a) + \mu_1 \phi(a_1) + \cdots + \mu_d \phi(a_d) = 0$$

for any $\phi \in \Phi$. If μ were equal to zero then we could easily contradict (i) by setting $f_i = 1$ and $f_j = 0$ for $j \neq i$. Therefore this is not the case and we may divide (5) through by μ to get (ii) (with $\lambda_i = -\mu^{-1} \mu_i$ for $i = 1, \ldots, d$). If the constant map lies in Φ then (5) must hold for it, which implies that $\lambda_1 + \cdots + \lambda_d = 1$. ∎

Lemma 13. *Let W and W' be finite-dimensional vector spaces over a field F and let $A \subseteq W$ be a finite non-empty set of Freiman dimension $r = r_F(A)$. Then there is a subset $\{a_1, \ldots, a_{r+1}\} \subseteq A$ with the following properties.*
(i) For any $w'_1, \ldots, w'_{r+1} \in W'$ there is a unique $\phi \in \operatorname{Hom}_2(A, W')$ with $\phi(a_i) =$

w_i' $(i=1,\ldots,r+1)$. *Moreover, if* $w_1',\ldots,w_{r+1}'$ *are linearly independent then* ϕ *is an isomorphism.*

(ii) *If* $a \in A$ *then there are* $\lambda_1,\ldots,\lambda_{r+1} \in F$ *(depending only on* a*) with* $\lambda_1+\cdots+\lambda_{r+1}=1_F$ *and such that* $\phi(a)=\lambda_1\phi(a_1)+\cdots+\lambda_{r+1}\phi(a_{r+1})$ *for any* $\phi \in \mathrm{Hom}_2(A,W')$.

Proof. Part (ii) and the first part of (i) are swift consequences of Lemma 12, (i). Although that lemma deals only with homomorphisms from A to F, the required extension follows easily using the isomorphism $\mathrm{Hom}_2(A,F^n) \cong (\mathrm{Hom}_2(A,F))^n$.

To prove the second assertion of (i), suppose that suppose that $\phi \in \mathrm{Hom}_2(A,W')$ is not a Freiman isomorphism. Then there are elements $a,a',a'',a''' \in A$ such that $\phi(a)+\phi(a')=\phi(a'')+\phi(a''')$ but $a+a' \neq a''+a'''$. By part (ii) there are $u_i,u_i',u_i'',u_i''' \in F$ such that $\psi(a)=\sum_{i=1}^{r+1} u_i\psi(a_i)$ for all $\psi \in \mathrm{Hom}_2(A,V)$, for any vector space V over F (and similarly for a',a'',a'''). One has

$$(6) \qquad \psi(a) + \psi(a') - \psi(a'') - \psi(a''') = \sum_{i=1}^{r+1}(u_i + u_i' - u_i'' - u_i''')\psi(a_i)$$

for all ψ. Taking $V=W$ and $\psi=\mathrm{id}$ we see that not all of the $u_i+u_i'-u_i''-u_i'''$ are zero. Therefore, setting $V=W'$ and $\psi=\phi$, one sees that $\phi(a_1),\ldots,\phi(a_{r+1})$ must be linearly dependent. ∎

Corollary 14. *Let* $A \subseteq W$. *Then* $r = r_F(A)$ *is the largest integer* d *such that* A *is isomorphic to a subset of* F^d *which is not contained in an affine subspace. In particular if* A *is a set of integers then* $r_{\mathbb{Q}}(A)=d$ *really is equal to dimension in the original sense of Freiman* [8].

Proof. Apply Lemma 13 (ii) with $W' = W$ and $\phi = \mathrm{id}$. It follows that there are $\{a_1,a_2,\ldots,a_{r+1}\}$ such that A is a subset of $\{\lambda_1 a_1+\cdots+\lambda_{r+1}a_{r+1}: \lambda_1+\cdots+\lambda_{r+1}=1_F\}$, an affine space over F of dimension r. As $r_F(A)$ is clearly an isomorphism invariant, the same must be true of any set isomorphic to A and it therefore follows that $d \leq r$. To see that $r \leq d$, pick linearly independent vectors $w_1',\ldots,w_{r+1}'$ in some space W', and observe that A is isomorphic to some set containing all of the w_i' by Lemma 13 (i). ∎

4. Isomorphism classes with small doubling

As in the last section, let W be a vector space over a field F. The main result of this section is a bound for the number of mutually non-isomorphic

sets $A \subseteq W$ of cardinality k with small (restricted) doubling property. The cardinality of $A \hat{+} A$ is an isomorphism invariant and we say that an isomorphism class of sets has small doubling property if $m = |A \hat{+} A|$ is relatively small in comparison to $|A|$ for each set in that class.

If $B \subseteq W$ and if t is a non-negative integer than a *t-extension* of B is simply any set $B \cup E$ where $E = \{x_1, \ldots, x_t\}$ and the x_i are distinct elements of $W \setminus B$.

Proposition 15. *Suppose that $A \subseteq W$ has sufficiently large cardinality $k = |A|$ and let $m = |A \hat{+} A|$. Then there exist $a^* \in A$ and $A_0, A_1 \subseteq A$ such that $|A_0| \leq 4k^{-1/15}m$, $|A \setminus A_1| \leq 4k^{-4/5}m$ and $a^* + A_1 \subseteq A_0 \hat{+} A_0$.*

Proof. Let $Q = \lfloor k^{1/5} \rfloor$ and $q = k^{-1/15}$. Draw a graph on vertex set A by joining x to y if $x+y$ is Q-*unpopular*, that is if the number of pairs $(w, z) \in A^2$, $w \neq z$, with $w+z = x+y$ is less than Q. The total number of edges in the graph is no more than Qm, and so some vertex a^* has degree at most $2Qm/k$. This means that with at most $2Qm/k$ exceptions the sums $a^* + a$, $a \in A$ are Q-popular. Pick a set $X \subseteq A$ by choosing each $a \in A$ to lie in X independently and at random with probability q. Let $Z = \{a \in A | a^* + a \notin X \hat{+} X\}$. If $s = a^* + a$ is Q-popular then there are at least $Q/3$ disjoint pairs $(b, c) \in A^2$ with $b + c = s$ and $b \neq c$. For each the probability that both b and c lie in X is q^2, and so

$$\mathbb{P}(a \in Z) \leq (1 - q^2)^{Q/3} \leq e^{-q^2 Q/3}.$$

It follows that

$$\mathbb{E}|Z| \leq \frac{2Qm}{k} + e^{-q^2 Q/3} k \leq 3k^{-4/5}m.$$

Clearly we have $\mathbb{E}|X| = qk = k^{14/15}$ and so

(7) $$\mathbb{E}\left(|X| + k^{11/15}|Z|\right) \leq 4k^{-1/15}m.$$

Pick a specific $X \subseteq A$ for which $|X| + k^{11/15}|Z| \leq 4k^{-1/15}m$, and set $A_0 = X$, $A_1 = A \setminus Z$. Then $|A_0| \leq 4k^{-1/15}m$, $|A \setminus A_1| \leq 4k^{-4/5}m$ and, by construction, $a^* + A_1 \subseteq A_0 \hat{+} A_0$. ∎

Lemma 16. *Fix a non-negative integer t and a set $B \subseteq W$ of cardinality $|B| = l$. Then the number of mutually non-isomorphic sets A of cardinality $|A| = l + t$, such that there exists a subset $A_0 \subseteq A$ satisfying $A_0 \cong_3 B$, is at most l^{3t^4}.*

Proof. Write $B = \{b_1, \ldots, b_l\}$. Given $A_0 \subseteq A$ with the required properties write $A_0 = \{a_1, \ldots, a_l\}$ and $A = A_0 \cup \{a_{l+1}, \ldots, a_{l+t}\}$ in a manner that respects the 3-isomorphism between A_0 and B; that is, for $i_1, i_2, i_3, i_4, i_5, i_6 \in \{1, \ldots, l\}$

$$b_{i_1} + b_{i_2} + b_{i_3} - b_{i_4} - b_{i_5} - b_{i_6} = 0$$

if and only if

$$a_{i_1} + a_{i_2} + a_{i_3} - a_{i_4} - a_{i_5} - a_{i_6} = 0.$$

The isomorphism class of A is determined by the 2-relations in V_{l+t} satisfied by A. With each possible relation we can associate its projection $w \in V_t$ onto the last t coordinates of V_{l+t}. There are at most t^4 non-zero projections w; fix one of them, say $w = (1, 0, \ldots, 0)$, and consider relations corresponding to this specific projection. If A satisfies such a relation, say

$$(8) \qquad\qquad a_{i_1} + a_{l+1} = a_{i_2} + a_{i_3}$$

with $i_1, i_2, i_3 \in \{1, \ldots, l\}$, then all other such relations are uniquely determined: they correspond to the equalities $a_{i_2} + a_{i_3} - a_{i_1} = a_{j_2} + a_{j_3} - a_{j_1}$ with $j_1, j_2, j_3 \in \{1, \ldots, l\}$, hence to the equalities $b_{i_2} + b_{i_3} - b_{i_1} = b_{j_2} + b_{j_3} - b_{j_1}$ with $j_1, j_2, j_3 \in \{1, \ldots, l\}$. Since there are at most l^3 choices for the relation (8), there are at most l^3 choices for the whole system of relations corresponding to w which A satisfies. Moreover, it is not difficult to see that this conclusion is actually true for any possible $w \neq 0$. The result follows on observing that to determine the system of relations in V_{l+t}, satisfied by A, we must for each non-zero $w \in V_t$ determine the system of relations corresponding to w. ∎

The next lemma is an easy exercise.

Lemma 17. *Suppose that $X \cong_6 X'$. Then $X \hat{+} X \cong_3 X' \hat{+} X'$ and any $B \subseteq X \hat{+} X$ is 3-isomorphic to a subset of $X' \hat{+} X'$.*

Proposition 18. *Let k be a sufficiently large positive integer and let $m \leq k^{31/30}$. Then the number of isomorphism classes of $A \subseteq F^k$ with $|A| = k$ and $|A \hat{+} A| \leq m$ is at most $(em/k)^k \exp\left(k^{31/32}\right)$.*

Proof. By Proposition 15, to any $A \subseteq F^k$ with $|A| = k$ and $|A \hat{+} A| \leq m$ there correspond $a^* \in A$ and $A_0, A_1 \subseteq A$ such that $|A_0| \leq 4k^{-1/15}m$, $|A_1| \geq |A| - 4k^{-4/5}m$ and $a^* + A_1 \subseteq A_0 \hat{+} A_0$. By Lemma 11, the number of possible 6-isomorphism classes to which A_0 can belong is $M < \exp\left(k^{30/31}\right)$; let $X_1, \ldots, X_M$ be representatives of these classes. By Lemma 17, the set $a^* + A_1$ is 3-isomorphic to a non-empty subset of some $X_i \hat{+} X_i$ of cardinality at most k; consequently, the number of 3-isomorphism classes to which A_1 can belong

is at most $Mk(em/k)^k$. By Lemma 16, each of these 3-isomorphism classes gives rise, by t-extensions for any fixed $t \le 4k^{-4/5}m$, to at most $\exp(k^{29/30})$ isomorphism classes. Altogether, we get at most

$$\exp(k^{30/31}) \cdot k(em/k)^k \cdot 4k^{-4/5}m \cdot \exp(k^{29/30}) < (em/k)^k \exp(k^{31/32})$$

possible isomorphism classes for A.   ■

5. Avoiding very small doubling

The bound of Proposition 18 becomes rather ineffective when $m \le 7k$, say. This section contains a trick which allows us to avoid ever considering sets with extremely small doubling (see, however, the section entitled Concluding Remarks). The result we will actually use, Corollary 20, is a consequence of the following more general proposition.

Proposition 19. *Let k be a sufficiently large positive integer and let A be a subset of an abelian group with $|A| = k$. Let $\epsilon > 0$ be sufficiently small. Then there is a set $B \subseteq A$ with*

$$|B| \le \frac{3\log(1/\epsilon)}{\epsilon}\sqrt{k}$$

and $|B \hat{+} B| \ge (1-\epsilon)k$.

Proof. Let $p = 3\epsilon^{-1}\sqrt{\log(6/\epsilon)/2k}$. Choose a set $B \subseteq A$ by picking each element of k independently and at random with probability p. By some rather standard tail estimates (see, for example, [4, Theorem A.1.4]) we have

$$(9) \qquad \mathbb{P}\left(||B| - pk| \ge \epsilon pk/3\right) < 2e^{-2\epsilon^2 p^2 k/9} = \epsilon/3.$$

Now for any $x \in B \hat{+} B$ write $s(x)$ for the number of ordered pairs $(b_1, b_2) \in B^2$ with $b_1 \ne b_2$ and $b_1 + b_2 = x$. The sum $M(B) = \sum_x s(x)^2$ equals the number of quadruples $(b_1, b_2, b_3, b_4) \in B^4$ with $b_1 + b_2 = b_3 + b_4$ and $b_1 \ne b_2$, $b_3 \ne b_4$. We call these summing quadruples.

Let us divide the summing quadruples $(a_1, a_2, a_3, a_4) \in A^4$ into two classes: those in which a_1, a_2, a_3, a_4 are all distinct and those in which either $a_1 = a_3, a_2 = a_4$ or $a_1 = a_4, a_2 = a_3$. Call this latter type degenerate; there are less than $2k^2$ degenerate quadruples in A^4, and at most k^3 non-degenerate ones. Now for a given quadruple, one can calculate the probability that it is also

contained in B^4. This probability is p^4 if the quadruple is non-degenerate, and p^2 otherwise. Thus

$$\mathbb{E}M(B) = \mathbb{E}\left(\sum s(x)^2\right) \leq 2k^2p^2 + k^3p^4.$$

It follows immediately that

$$(10) \qquad \mathbb{P}\left(M(B) \leq \frac{2k^2p^2 + k^3p^4}{1 - \epsilon/3}\right) \geq \epsilon/3.$$

Comparing (9) and (10), we see that there is a specific choice of B for which

$$(11) \qquad ||B| - pk| \leq \epsilon pk/3 \qquad \text{and} \qquad M(B) \leq \frac{2k^2p^2 + k^3p^4}{1 - \epsilon/3}.$$

Now the Cauchy–Schwarz inequality gives

$$M(B) = \sum_{x \in B\hat{+}B} s(x)^2 \geq \frac{\left(\sum s(x)\right)^2}{|B\hat{+}B|} = \frac{|B|^2(|B| - 1)^2}{|B\hat{+}B|}.$$

This, together with (11) and a small amount of computation, shows that $|B\hat{+}B| \geq (1-\epsilon)k$. Moreover we have, from (11), that

$$|B| \leq pk(1 + \epsilon/3) \leq \frac{3\log(1/\epsilon)}{\epsilon}\sqrt{k}.$$

This completes the proof. ∎

Corollary 20. *Let k be a sufficiently large positive integer and let A be a subset of an abelian group with $|A| = k$. Then A contains a subset C, $|C| \geq k/8$, with $|C\hat{+}C| \geq 7|C|$.*

Proof. Apply Proposition 19 with $\epsilon = 1/9$. This supplies us with a set B, $|B| \leq 100\sqrt{k}$, such that $|B\hat{+}B| \geq 8k/9$. Now simply take C to be the union of B and some arbitrary elements of A so as to ensure that $|C|$ lies between $k/8$ and $8k/63$. ∎

6. The rank of sets with small doubling

In this section we obtain results concerning the Freiman dimension of sets $A \subseteq \Gamma$ with $|A| = k$ and $|A\hat{+}A| = m$. It is in this section that the difference between different groups Γ comes to the fore. When $\Gamma = \mathbb{Z}_N$ we can relate A to a subset of the integers, but this is not possible for $\Gamma = \mathbb{Z}_2^n$ and, by necessity, we must find a different approach.

Let $A \subseteq \mathbb{Z}_N$ have $|A| = k$ and $|A \hat{+} A| = m$ and consider the image $\overline{A} \subseteq \mathbb{Z}$ of A under the least positive residue or "unfolding" map $\mathbb{Z}_N \hookrightarrow \{1, \dots, N\}$. This clearly has $|\overline{A}| = k$ and $|\overline{A} \hat{+} \overline{A}| \leq 2m$. Thus we may apply the following result of Freiman [8, page 24].

Lemma 21 (Freiman). *Let $S \subseteq \mathbb{Z}$ have cardinality k and Freiman dimension $r = r_{\mathbb{Q}}(S)$. Then we have the inequality*

$$|S \hat{+} S| \geq r \left(k - \frac{r+1}{2} \right).$$

Lemma 22. *Suppose that $A \subseteq \mathbb{Z}_N$ has $|A| = k$ and $|A \hat{+} A| = m$, and let $\overline{A}$ be the image of A under the unfolding map $\mathbb{Z}_N \hookrightarrow \mathbb{Z}$. Then we have the bound $r_{\mathbb{Q}}(\overline{A}) \leq 4m/k$.*

Proof. Observe that $r = r(\overline{A}) \leq k - 1$, this being a general property shared by all sets of k integers. Applying Lemma 21, we have

$$2m \geq |\overline{A} \hat{+} \overline{A}| \geq r \left(k - \frac{r+1}{2} \right) \geq \frac{1}{2} kr. \qquad \blacksquare$$

Proposition 23. *We have the bounds*

$$(12) \qquad\qquad |S_k^m(\mathbb{Z}_N)| \leq N^{1 + \frac{4m}{k}} \left(\frac{2em}{k} \right)^k \exp \left(k^{31/32} \right)$$

if $m \leq k^{31/30}/2$ and

$$(13) \qquad\qquad |S_k^m(\mathbb{Z}_N)| \leq N^{1 + \frac{4m}{k}} k^{4k}$$

whatever the value of m.

Proof. It suffices to count the number of lifts $\overline{A}$ of sets $A \in S_k^m(\mathbb{Z}_N)$. Let $X_1, \dots, X_M$ be a complete set of representatives of the isomorphism classes of subsets X of the integers with $|X| = k$ and $|X \hat{+} X| \leq 2m$. We always have $M \leq k^{4k}$ by Lemma 11 but, when $m \leq k^{31/30}/2$, Proposition 18 allows us to do better and we have the bound $M \leq (2em/k)^k \exp(k^{31/32})$. Now each $\overline{A}$ is the image of some X_i under some isomorphism $\psi : X_i \to \{1, \dots, N\}$, and by Lemma 13 (i) there are at most $N^{r_{\mathbb{Q}}(X_i)+1}$ such isomorphisms. The proposition follows immediately from these observations and Lemma 22. $\blacksquare$

We now turn our attention to subsets of $\mathbb{Z}_2^n$. An important tool will be the following consequence of an inequality of Plünnecke which was reproved by Ruzsa [14]. If G is any abelian group and if $X \subseteq G$ then we write kX for the k-fold sum $X + X + \cdots + X$ and $kX - lX$ for the set $X + \cdots + X - X - \cdots - X$, where there are $k-1$ plus signs and l difference signs.

Theorem 24 (Plünnecke–Ruzsa). *Let $A\subseteq G$ be such that $|A+A|\leq C|A|$. Then for any positive integers k and l one has*

$$|kA - lA| \leq C^{k+l}|A|.$$

This allows us to get a bound on the Freiman dimension over $\mathbb{Z}_2$ of subsets of $\mathbb{Z}_2^n$ with small doubling.

Lemma 25. *Suppose that $A\subseteq\mathbb{Z}_2^n$ has cardinality k and $|A\hat{+}A|=m$. Then $r_{\mathbb{Z}_2}(A)+1\leq 4m\log k/k$.*

Proof. By Corollary 14 we may assume (by subjecting A to a Freiman isomorphism if necessary) that A is not contained in any affine subspace of $\mathbb{Z}_2^n$ of dimension less than $r_{\mathbb{Z}_2}(A)$.

Let $\prec$ be an arbitrary ordering on $\mathbb{Z}_2^n$ and let $h=\lfloor\log_e k\rfloor$. Suppose that $|A|=k$ and that $|A\hat{+}A|=m$, so that $|A+A|\leq m+1$ (we are in characteristic 2!), and let $X\subseteq A$ be a subset of A which is such that the sums $x_1+x_2+\cdots+x_h$ $(x_1\prec\cdots\prec x_h)$ are all distinct, and which is maximal with respect to this property. If $|X|\geq h$ then we have, using Theorem 24,

$$\left(\frac{|X|}{h}\right)^h \leq \binom{|X|}{h}$$
$$\leq |hX|$$
$$\leq |hA|$$
$$\leq \left(\frac{m+1}{k}\right)^h k.$$

It follows immediately from this that

(14) $$|X| \leq 2mhk^{-1+1/h} \leq 4m\log k/k.$$

Now the maximality of X implies that if we set $Y=X\cup\{a\}$ for any $a\in A\backslash X$ then some two of the sums $y_1+\cdots+y_h$, $y_1'+\cdots+y_h'$ must be equal, where $y_1\prec\cdots\prec y_h$ and $y_1'\prec\cdots\prec y_h'$. It is clear that precisely one of the y_i's or y_i''s must equal a, and so $a\in hX-(h-1)X$. In particular, A is contained in the affine subspace of $\mathbb{Z}_2^n$ spanned by X. The required inequality follows. ∎

It is interesting to compare this with the estimate

$$r_{\mathbb{Z}_2}(A) + 1 \leq \log k + (m/k)^4 + 2\log(m/k).$$

This bound, which follows from the work of Ruzsa [16] on Freiman's theorem in torsion groups, is stronger than Lemma 25 when $m/k\ll(\log k)^{1/3}$. Ruzsa and I [11], can improve this bound to something more like $\log k+2(m/k)^2$. It would be interesting to know the true behaviour: perhaps $r\leq\log k+Cm/k$.

Proposition 26. *We have the bounds*

(15)
$$|S_k^m(\mathbb{Z}_2^n)| \leq N^{\frac{4m \log k}{k}} \left(\frac{em}{k}\right)^k \exp\left(k^{31/32}\right)$$

if $m \leq k^{31/30}$ *and*

(16)
$$|S_k^m(\mathbb{Z}_2^n)| \leq N^{\frac{4m \log k}{k}} k^{4k}$$

whetever the value of m.

Proof. This is slightly more straightforward than Proposition 23. Let $X_1, \ldots, X_M$ be a complete set of representatives of the isomorphism classes of sets $X \subseteq \mathbb{Z}_2^n$ with $|X| = k$ and $|X \hat{+} X| \leq m$. Proposition 18 gives bounds for M, and we know that any set $A \in S_k^m(\mathbb{Z}_2^n)$ is the image of some X_i under an isomorphism $\psi : X_i \to \mathbb{Z}_2^n$. Lemma 13, (i) and Lemma 25 allow us to bound the number of such isomorphisms. ∎

7. Putting everything together

In this section we use Propositions 23 and 26 to give upper bounds for the sums
$$\sum_{m \geq 7k} |S_k^m(\Gamma)| 2^{-m}$$
when $\Gamma = \mathbb{Z}_N$ and $\mathbb{Z}_2^n$. An application of Proposition 19 will then allow us to prove Theorems 7 and 9. We begin with Theorem 7. Let $k = \lfloor 20 \log N \rfloor$. Using (12) and (13) we have

$$\sum_{m \geq 7k} |S_k^m(\mathbb{Z}_N)| 2^{-m} = \sum_{m=7k}^{\lfloor k^{31/30}/2 \rfloor} |S_k^m(\mathbb{Z}_N)| 2^{-m} + \sum_{m=\lfloor k^{31/30}/2 \rfloor}^{k(k-1)/2} |S_k^m(\mathbb{Z}_N)| 2^{-m}$$

$$\leq \sum_{m=7k}^{\lfloor k^{31/30}/2 \rfloor} 2^{m\left(\left(\frac{4}{k}+\frac{1}{m}\right)\log N + \frac{k}{m}\log\left(\frac{2em}{k}\right) - 1 + o(1)\right)}$$

$$+ \sum_{m \geq \lfloor k^{31/30}/2 \rfloor} 2^{m\left(\frac{4\log N}{k} - 1 + o(1)\right)}.$$

Now $\log(2eC)/C - 1 \leq -0.2499$ when $C \geq 7$, and so both of the sums appearing here are bounded above by N^{-2} for N sufficiently large.

Now suppose that some randomly selected set $A \subseteq \mathbb{Z}_N$ contains $X \hat{+} X$ with $|X| = k$. Then, by Proposition 19, A contains some set $Y \hat{+} Y$ where

$|Y| \geq k/8$ and $|Y \hat{+} Y| \geq 7|Y|$. However if $k = \lceil 160 \log N \rceil$ then the expected number of such sets Y that A contains is at most N^{-2}, and in particular the probability that it contains even one is $o(1)$. It follows immediately that almost surely A does not contain any $X \hat{+} X$ with $|X| = \lceil 160 \log N \rceil$, and this is the claim of Theorem 7. ∎

Similar computations apply for the group $\mathbb{Z}_2^n$, but now of course we use formulae (15) and (16) of Proposition 26 instead. Let $k = \lfloor 11 \log N \log \log N \rfloor$. One has

$$\sum_{m \geq 7k} |S_k^m(\mathbb{Z}_2^n)| 2^{-m} = \sum_{m=7k}^{\lfloor k^{31/30}/2 \rfloor} |S_k^m(\mathbb{Z}_2^n)| 2^{-m} + \sum_{m=\lfloor k^{31/30}/2 \rfloor}^{k(k-1)/2} |S_k^m(\mathbb{Z}_2^n)| 2^{-m}$$

$$\leq \sum_{m=7k}^{\lfloor k^{31/30}/2 \rfloor} 2^{m\left(\frac{4 \log k \log N}{k} + \frac{k}{m} \log\left(\frac{em}{k}\right) - 1 + o(1)\right)}$$

$$+ \sum_{m \geq \lfloor k^{31/30}/2 \rfloor} 2^{m\left(\frac{4 \log k \log N}{k} - 1 + o(1)\right)}.$$

Now $\log(eC)/C - 1 \leq -0.39$ when $C \geq 7$, and so both sums appearing here are bounded above by N^{-2} for N sufficiently large. The rest of the argument proceeds exactly as before, and one gets Theorem 9 (with $C_2 = 308$). ∎

8. A lower bound for $\mathbb{Z}_2^n$

In this section we prove the lower bound in Theorem 9. Throughout this section we write $\Gamma = \mathbb{Z}_2^n$ and $N = |\Gamma|$.

Proposition 27.

$$\lim_{n \to \infty} \mathbb{P}_\Gamma(\omega \geq \tfrac{1}{4} \log N \log \log N) = 1.$$

Proof. Let $k = \lfloor \log n + \log \log n - 1 \rfloor$ and consider all k-dimensional subspaces $H_1, \ldots, H_M$ of Γ, where

$$M = \frac{(2^n - 1)(2^n - 2) \cdots (2^n - 2^{k-1})}{(2^k - 1)(2^k - 2) \cdots (2^k - 2^{k-1})} \geq 2^{nk-k^2}.$$

For a random set $A \subseteq \Gamma$ let X_i denote the indicator function of the event $H_i \setminus \{0\} \subseteq A$, and let $X = X_1 + \cdots + X_M$. We have

(17) $\mathbb{E}X \geq 2^{nk-k^2-2^k}.$

We will also estimate $\mathrm{Var}(X)$, and to do this it is necessary to prove a lemma.

Lemma 28. *The number of pairs (i,j) for which $\dim(H_i \cap H_j) = l$, $0 \leq l \leq k$, is at most $2^{n(2k-l)}$.*

Proof. We will show that the quantity in question is exactly

$$\frac{(2^n - 1)(2^n - 2) \cdots (2^n - 2^{2k-l-1})}{(2^k - 2^l)^2 \cdots (2^k - 2^{k-1})^2 (2^l - 1)(2^l - 2) \cdots (2^l - 2^{l-1})},$$

which is easily seen to be at most $2^{n(2k-l)}$. To make this count, first pick the intersection $L = H_i \cap H_j$ in any of

$$\frac{(2^n - 1)(2^n - 2) \cdots (2^n - 2^{l-1})}{(2^l - 1)(2^l - 2) \cdots (2^l - 2^{l-1})}$$

ways. Then pick vectors $u_1, \ldots, u_{k-l}, v_1, \ldots, v_{k-l}$ so that $H_i = \mathrm{Span}(L, u_1, \ldots, u_{k-l})$ and $H_j = \mathrm{Span}(L, v_1, \ldots, v_{k-l})$. The subspace L together with all the points u_i, v_i spans a subspace of dimension $2k - l$, so the number of choices for the u_i, v_i is

$$(2^n - 2^l)(2^n - 2^{l+1}) \cdots (2^n - 2^{2k-l-1}).$$

However, many of these choices will give the same subspaces H_i, H_j. If H_i, H_j are fixed and have $H_i \cap H_j = L$ then there are $(2^k - 2^l) \cdots (2^k - 2^{k-1})$ ways of choosing $u_1, \ldots, u_{k-l}$ so that $H_i = \mathrm{Span}(L, u_1, \ldots, u_{k-l})$, and the same number of ways of choosing $v_1, \ldots, v_{k-l}$. This completes the proof of the lemma. ∎

Now the covariance $\mathrm{Cov}(X_i, X_j)$ is positive if and only if $H_i \backslash \{0\}$ intersects $H_j \backslash \{0\}$, which requires $\dim(H_i \cap H_j) \geq 1$. Thus we have, using the lemma,

$$\mathrm{Var}(X) = \sum_{i,j} \mathrm{Cov}(X_i, X_j)$$

$$\leq 2 \sum_{l=1}^{k} 2^{n(2k-l)} 2^{-2^{k+1}} \left(2^{2^l} - 1 \right)$$

$$\leq 2 \sum_{l=1}^{k} 2^{2kn - 2^{k+1} + 2^l - nl}.$$

Now one can check that $2^l - nl \leq 2 - n$ for $1 \leq l \leq k$, and so

$$\mathrm{Var}(X) \leq 8k 2^{(2k-1)n - 2^{k+1}}.$$

Now Chebyshev's inequality tells us that

$$\mathbb{P}(X = 0) \leq \frac{\mathrm{Var}(X)}{(\mathbb{E}X)^2} \leq 8k 2^{2k^2 - n},$$

which is $o(1)$. This means that with probability $1 - o(1)$ the set A contains some $H_i \backslash \{0\}$. Note, however, that $H_i \backslash \{0\}$ is precisely $H_i \hat{+} H_i$. ∎

9. Concluding remarks

It would certainly be of interest to determine the clique number of a random Cayley sum graph on $\mathbb{Z}_N$ asymptotically (almost surely). There are several points at which are argument leading to Theorem 7 was suboptimal. One was that we rather shied away from the issue of counting sets with extremely small doubling property, resorting instead to the trick of Proposition 19. It is possible to say rather more about these using Freiman's deep theorem [8], which says that such sets are economically contained in a multidimensional arithmetic progression. It was also rather wasteful to treat subsets of $\mathbb{Z}_N$ by using the unfolding map $\mathbb{Z}_N \hookrightarrow \{1,\ldots,N\}$ in such a naïve way. When N is prime, it is possible to prove the following result:

Proposition 29. *Suppose that N is prime and that A is a subset of $\mathbb{Z}_N$ with $|A \hat{+} A| \leq (\log N)^2/32$. Then there are $\lambda \in \mathbb{Z}_N^*, \mu \in \mathbb{Z}_N$ for which $\lambda A + \mu \subseteq \{0,\ldots,\lfloor N/2 \rfloor\}$.*

By subjecting our set A to such a homothety one can ensure that the unfolding map induces a Freiman isomorphism, so that the lifting $\overline{A}$ has $|\overline{A} \hat{+} \overline{A}| = m$. Proposition 29 may be called a *rectification principle* in the spirit of [5].

Using these results, and tidying up the approximations used in, for example, Lemma 22, one can certainly reduce the number 160 drastically, and perhaps even to $(3+\epsilon)$. To do this would be difficult, and there seem to be some even more substantial obstacles to pushing the constant down to $(2+\epsilon)$, which seems likely to be the correct value. We may return to these issues in a future paper.

Another issue deserving of remark arises from the result of Graham and Ringrose [9], which implies that there is an absolute constant c such that, for infinitely many primes N, the clique number $\omega(P_N)$ is at least $c \log N \log\log\log N$. However, using the Borel-Cantelli lemma together with the bounds obtained in §7 one can show that almost surely any sequence $(G_N)_N$ of random Cayley sum graphs on $\mathbb{Z}_N$, $|G_N| = N$, will have $\limsup_{N \to \infty} \omega(G_N)/\log N \leq 160$. Thus Paley graphs are (in a sense) not as good Ramsey graphs as random Cayley sum graphs.

Let us close by mentioning a conjecture from [2].

Conjecture 30 (Alon). There exists an absolute constant b such that for every group Γ on n elements there is a Cayley graph on this group containing neither a complete subgraph nor an independent set on more than $b \log n$ vertices.

The results of this paper confirm this conjecture for cyclic groups but one should not expect the full conjecture, if it is true, to be easy to prove. Quite apart from the fact that our methods seem to break down hopelessly for all but a few very special groups, the result of Proposition 27 shows that random methods alone will not suffice to attack Alon's conjecture.

10. Acknowledgements

References

[1] P. K. AGARWAL, N. ALON, B. ARONOV and S. SURI: Can Visibility Graphs be Represented Compactly?, *Disc. Comput. Geom.* **12** (1994), 347–365.

[2] N. ALON: Graph powers, *Contemporary combinatorics, Bolyai Soc. Math. Stud.* **10** (2002), 11–28, János Bolyai Math. Soc., Budapest.

[3] N. ALON and A. ORLITSKY: Repeated Communication and Ramsey Graphs, *IEEE Transactions on Information Theory* **41** (1995).

[4] N. ALON and J. SPENCER: *The probabilistic method*, 2nd Ed., Wiley, 2000.

[5] Y. F. BILU, V. F. LEV and I. Z. RUZSA: Rectification principles in additive number theory, *Discrete Comput. Geom.* **19(3)** Special Issue (1998), 343–353.

[6] B. BOLLOBÁS: Modern Graph Theory, *Graduate Texts in Mathematics* **184** (1998), Springer.

[7] S. L. G. CHOI: On irregularities of integer sequences, *Mathematika* **21** (1974), 101–106.

[8] G. A. FREIMAN: *Foundations of a structural theory of set addition*, Translations of Mathematical Monographs, **Vol. 37**, American Mathematical Society, Providence, R. I., 1973. vii+108 pp.

[9] S. W. GRAHAM and C. J. RINGROSE: Lower bounds for least quadratic nonresidues, *Analytic number theory* (Allerton Park, IL, 1989), 269–309, *Progr. Math.* **85** (1990), Birkhäuser Boston, Boston, MA.

[10] B. J. GREEN: *Lower bounds for random trigonometric sums*, in preparation.

[11] B. J. GREEN and I. Z. RUZSA: *Sets with small sumset and rectification*, submitted.

[12] V. F. LEV and S. V. KONYAGIN: Combinatorics and linear algebra of Freiman's isomorphism, *Mathematika* **47(1–2)** (2000), 39–51 (2002).

[13] M. KRIVELEVICH, B. SUDAKOV, V. H. VU and N. C. WORMALD: Random Regular Graphs of High Degree, *Random Structures Algorithms* **18(4)** (2001), 346–363.

[14] I. Z. RUZSA: Sums of Finite Sets, *Number Theory* (New York, 1991–1995), 281–293, Springer, New York, 1996.

[15] I. Z. RUZSA: *Uniform distribution, positive trigonometric polynomials and difference sets*, Seminar on Number Theory, 1981/1982, Exp. No. 18, Univ. Bordeaux I, Talence, 1982.

[16] I. Z. RUZSA: An analog of Freiman's theorem in groups, *Astérisque* **258** (1999), 323–326.

Ben Green

Trinity College
Cambridge CB2 1TQ
England
bjg23@hermes.cam.ac.uk