the color classes contains a proper generalized arithmetic progression $a + [0, n-1]^r \cdot (v_1, \ldots, v_r)$, where $a \in [0, n-1]^d$ and $v_1, \ldots, v_r \in [0, 1]^d$. (Hint: apply Theorem 6.15 with n replaced by n^r.)

6.3.12 (Gallai's theorem) Let $k \geq 1$, $d \geq 1$, $m \geq 1$, and let $v_1, \ldots, v_k$ be elements of $\mathbf{Z}^d$. Show that there exists an $N = N(k, d, m, v_1, \ldots, v_k)$ such that for partition of the cube $[1, N]^d \subset \mathbf{Z}^d$ into m color classes $E_1, \ldots, E_m$, then at least one of the color classes contains a set of the form $\{x + rv_1, \ldots, x + rv_k\}$ for some $x \in \mathbf{Z}^d$ and some non-zero integer r.

6.4 Proof of the Balog–Szemerédi–Gowers theorem

Let A and B be two additive sets with common ambient group. Let $G = G(A, B, E)$ be a bipartite graph whose color classes are A and B and whose edge set is E (an edge is a pair (a, b) where $a \in A$ and $b \in B$). Recall that the *partial sum set* $A \overset{G}{+} B$ is defined as the collection of the sums $a + b$ where $a \in A, b \in B$ and $(a, b) \in E$.

Balog and Szemerédi [16] proved that if A and B are two sets of cardinality N and $|E| \geq n^2/K$ and $|A \overset{G}{+} B| \leq K'n$ for some K, K', then one can find $A' \subset A$ and $B' \subset B$ such that $|A'|, |B'|, |A' + B'| = \Theta_{K,K'}(n)$.

As stated, the above theorem is only useful if K and K' are independent of n (or extremely slowly growing in n). With a new proof, Gowers [138] has recently strengthened this statement by showing that the implicit constants in the $\Theta_{K,K'}()$ notation can be taken to be polynomial in K and K', and hence the theorem remains effective even when K and K' are as large as n^ε for some absolute constant $\varepsilon > 0$; we have already stated this result in Theorem 2.29. This has proven to be immensely valuable in a number of applications in which polynomial-type bounds are desired, for instance in Gowers' proof of Szemerédi's theorem (see in particular Section 11.3). The polynomials in Gowers' proof were implicit, but by following his ideas, one can work out the explicit version given in Theorem 2.29. Our treatment here is based on that in [340].

As it turns out, one can view the Balog–Szemerédi–Gowers theorem as a statement about dense bipartite graphs. Clearly, if a bipartite graph $G(A, B, E)$ has many edges, then there will be many pairs of vertices $a \in A$, $b \in B$ which are connected by paths of length 1. One then expects there to be many pairs $a, a' \in A$ which are connected by paths of length two, and many pairs $a \in A, b \in B$ which are connected by paths of length three. Furthermore, this connectivity becomes increasingly more "uniform" as the length of the path increases; compare with the results on arithmetic progressions in sum sets in Section 4.7. It is this uniformity which is essential to the proof of the Balog–Szemerédi–Gowers theorem.

We begin by formalizing the above principle for paths of length two and length three.

Lemma 6.19 (Paths of length two) *Let $G(A, B, E)$ be a bipartite graph with $|E| \geq |A||B|/K$ for some $K \geq 1$. Then, for any $0 < \varepsilon < 1$, there exists a subset $A' \subseteq A$ such that*

$$|A'| \geq \frac{|A|}{\sqrt{2}K}$$

and such that at least $(1 - \varepsilon)$ of the pairs of vertices $a, a' \in A'$ are connected by at least $\frac{\varepsilon}{2K^2}|B|$ paths of length two in G.

Proof By decreasing K if necessary we may assume $|E| = |A||B|/K$. Observe the combinatorial identities

$$\mathbf{E}_{b\in B}\frac{|N(b)|}{|A|} = \mathbf{E}_{a\in A}\frac{|N(a)|}{|B|} = \frac{|E|}{|A||B|} = \frac{1}{K}$$

and

$$\mathbf{E}_{b\in B}\frac{|N(b)|^2}{|A|^2} = \mathbf{E}_{a,a'\in A}\frac{|N(a)\cap N(a')|}{|B|}.$$

Applying Cauchy–Schwarz we conclude that

$$\mathbf{E}_{a,a'\in A}\frac{|N(a)\cap N(a')|}{|B|} \geq \frac{1}{K^2}.$$

Let Ω be the set of all pairs (a, a') such that $|N(a) \cap N(a')| < \frac{\varepsilon}{2K^2}|B|$; in other words, $(a, a') \in \Omega$ if a, a' are *not* connected by at least $\frac{\varepsilon}{2K^2}$ paths of length two. Clearly we have

$$\mathbf{E}_{a,a'\in A}\mathbf{I}((a, a') \in \Omega)\frac{|N(a)\cap N(a')|}{|B|} < \frac{\varepsilon}{2K^2}$$

and hence

$$\mathbf{E}_{a,a'\in A}\left(1 - \frac{1}{\varepsilon}\mathbf{I}((a, a') \in \Omega)\right)\frac{|N(a)\cap N(a')|}{|B|} \geq \frac{1}{2K^2}.$$

The left-hand side can be rearranged as

$$\mathbf{E}_{b\in B}\frac{1}{|A|^2}\sum_{a,a'\in N(b)}\left(1 - \frac{1}{\varepsilon}\mathbf{I}((a, a') \in \Omega)\right)$$

and hence by the pigeonhole principle there exists $b \in B$ such that

$$\frac{1}{|A|^2} \sum_{a,a' \in N(b)} \left(1 - \frac{1}{\varepsilon}\mathbf{I}((a,a') \in \Omega)\right) \geq \frac{1}{2K^2}.$$

In particular this implies that $|N(b)| \geq \frac{|A|}{\sqrt{2}K}$ and that $|\{a, a' \in N(b) : (a,a') \in \Omega)\}| \leq \varepsilon |N(b)|^2$. The claim then follows by setting $A' := N(b)$. □

We now obtain an analogous result for paths of length three.

Corollary 6.20 (Paths of length three) *Let $G(A, B, E)$ be a bipartite graph with $|E| \geq |A||B|/K$ for some $K \geq 1$. Then there exists $A' \subseteq A$, $B' \subseteq B$ with $|A'| \geq \frac{|A|}{4\sqrt{2}K}$ and $|B'| \geq \frac{|B|}{4K}$, such that every $a \in A'$ and $b \in B'$ is connected by at least $\frac{|A||B|}{2^{12}K^4}$ paths of length three.*

Proof Before we apply Lemma 6.19 it is convenient to prepare the graph G a little bit. Let $\tilde{A}$ be the set of vertices in A that have degree at least $|B|/2K$, and let $\tilde{G} = \tilde{G}(\tilde{A}, B, \tilde{E})$ be the induced subgraph. Since at most $|A||B|/2K$ edges are removed when passing from G to $\tilde{G}$, we see that $\tilde{G}$ has at least $|A||B|/2K$ edges. Writing $|A| = L|\tilde{A}|$ for some $L \geq 1$ and applying Lemma 6.19 to $\tilde{G}$ (with K replaced by $2K/L$ and $\varepsilon := \frac{1}{16K}$) we can find a subset $\tilde{A}'$ of A' of size

$$|\tilde{A}'| \geq \frac{|\tilde{A}|}{\sqrt{2}(2K/L)} = \frac{|A|}{2\sqrt{2}K}$$

and such that $1 - \frac{1}{16K}$ of the pairs $a, a' \in \tilde{A}'$ are connected by at least $L^2|B|/128K^3$ paths of length two.

Let us call a pair $(a, a') \in \tilde{A}' \times \tilde{A}'$ *bad* if they are not connected by at least $\frac{L^2|B|}{8K^2}$ paths of length two; thus there are at most $\frac{1}{16K}|\tilde{A}'|^2$ bad pairs. Let A' be the set of all $a \in \tilde{A}'$ such that at most $\frac{1}{8K}|\tilde{A}'|$ pairs (a, a') are bad. Then $|\tilde{A}' \backslash A'| \leq \frac{|\tilde{A}'|}{2}$, and thus

$$|A'| \geq \frac{1}{2}|\tilde{A}'| \geq \frac{|A|}{4\sqrt{2}K}.$$

Having constructed A', we turn now to B'. Since every element in $\tilde{A}$ (and hence in $\tilde{A}'$) has degree at least $|B|/2K$, we have

$$\sum_{b \in B} |\{a \in \tilde{A}' : (a,b) \in E)\}| = |\{(a,b) \in E : a \in A'\}| \geq |\tilde{A}'|\frac{|B|}{2K},$$

so if we let

$$B' := \left\{ b \in B : |\{a \in \tilde{A}' : (a,b) \in E)\}| \geq \frac{|\tilde{A}'|}{4K} \right\}$$

then we have

$$|\tilde{A}'||B'| \geq \sum_{b \in B'} |\{a \in \tilde{A}' : (a,b) \in E)\}| \geq |\tilde{A}'|\frac{|B|}{2K} - \frac{|\tilde{A}'|}{4K}|B| = \frac{|\tilde{A}'||B|}{4K}.$$

In particular we have $|B'| \geq |B|/4K$.

Finally, let $a \in A'$ and $b \in B'$ be arbitrary. By the construction of B', then b is adjacent to at least $|\tilde{A}'|/4K$ elements a' of $\tilde{A}'$. By construction of A', at most $|\tilde{A}'|/8K$ of the pairs (a, a') are bad. Thus there are at least $|\tilde{A}'|/8K \geq |A|/16\sqrt{2}K$ vertices a' which are simultaneously adjacent to b, and are connected to a by at least $\frac{L^2|B|}{8K^2}$ paths of length two. Thus a and b are connected by at least

$$\frac{|A|}{16\sqrt{2}K}\frac{L^2|B|}{128K^3} \geq \frac{|A||B|}{2^{12}K^4}$$

paths of length three. □

We can now derive as a consequence the Balog–Szemerédi–Gowers theorem, Theorem 2.29.

Proof of Theorem 2.29 First observe that we may ensure that A and B are disjoint, by the artificial trick of replacing the ambient group Z with $Z \times \mathbf{Z}$, replacing A with $A \times \{0\}$, and B with $B \times \{1\}$. Let us view the set $G \subset A \times B$ in the theorem as a bipartite graph on A and B. Applying Corollary 6.20, we can find A', B' obeying (2.18), (2.19), and such that every pair $a \in A'$, $b \in B'$ is connected by at least $|A||B|/2^{12}K^4$ paths of length three:

$$|\{(a', b') \in A \times B : (a, b'), (a', b'), (a', b) \in G\}| \geq \frac{|A||B|}{2^{12}K^4}.$$

Exploiting the obvious identity

$$a + b = (a + b') - (a' + b') + (a' + b)$$

and writing $x := a + b'$, $y := a' + b'$, $z := a' + b$, we conclude that

$$|\{(x, y, z) \in A \overset{G}{+} B : x - y + z = a + b\}| \geq \frac{|A||B|}{2^{12}K^4}.$$

Since the total number of triples (x, y, z) is at most

$$|A \overset{G}{+} B|^3 \leq (K')^3|A|^{3/2}|B|^{3/2},$$

we conclude that the total number of possible values for $a + b$ is at most $2^{12}K^4(K')^3|A|^{1/2}|B|^{1/2}$, and the claim follows. □

Note that in this proof it is not critical that the group is abelian. For a multiplicative group, we can replace $a + b = (a + b') - (a' + b') + (a' + b)$ by $ab = (ab')(a'b')^{-1}(a'b)$, and the rest of the proof is the same.

To conclude this section, let us mention a generalization of Balog–Szemerédi–Gowers result for hypergraphs. Let $A_1, \ldots, A_k$ be additive sets with common ambient group (which we may take to be disjoint, by the trick used above) and let E be some family of ordered k-tuples $(a_1, \ldots, a_k)$ such that $a_i \in A_i$, $1 \leq i \leq k$. The sets $A_1, \ldots, A_k$ together with E are known as a *k-uniform k-partite hypergraph* which we shall call H; the set E is then known as the *edge set* of H (notice that a bipartite graph is a special case when $k = 2$). We denote by $\bigoplus_{H i=1}^{k} A_i$ the collection of the sums $a_1 + \cdots + a_k$ where $(a_1, \ldots, a_k) \in E$. For the case $k = 2$, we are talking about bipartite graphs.

Theorem 6.21 *[340] Let $k \geq 1$, and let n, K be positive numbers. If $A_1, \ldots, A_k$ are additive sets in a group Z of cardinality at most n, then $H(A_1, \ldots, A_k, E)$ is a k-partite k-uniform hypergraph with at least n^k/K edges and $\left|\bigoplus_{H i=1}^{k} A_i\right| \leq Kn$, then one can find subsets $A'_i \subset A_i$ such that*

- $|A'_i| = \Omega_k(n/K^{O_k(1)})$ *for all* $1 \leq i \leq k$.
- $|A'_1 + \cdots + A'_k| = \Omega_k(K^{O_k(1)}n)$.

The heart of the proof is the following claim.

Claim 6.22 *Let $A_1, \ldots, A_k$ and n, K be as in the theorem above. Set $X = \bigoplus_{H i=1}^{k} A_i$. There are subsets $A'_i \subset A_i, i = 1, \ldots, k$ of cardinality at least $\Omega_k(n/K^{O_k(1)})$ and sets $Y_j \subseteq Z, 1 \leq j \leq 2k-2$ of cardinality at most $O_k(K^{O_k(1)}n)$, such that every element in $A'_1 + \cdots + A'_k$ can be written in the form $x + \sum_{j=1}^{2k-2} y_j$ where $x \in X, y_j \in Y_j$ in at least $\Omega_k(n^{2k-2}/K^{O_k(1)})$ ways.*

It is easy to deduce Theorem 6.21 from this claim. For the sets $A'_1, \ldots, A'_k$ as in the claim, we have

$$\begin{aligned} |A'_1 + \cdots + A'_k| &\leq \frac{|X| \prod_{j=1}^{2k-2} |Y_j|}{\Omega_k(n^{2k-2}/K^{O_k(1)})} \\ &= \Omega_k(K^{O_k(1)}n) \end{aligned}$$

as desired. The proof of Claim 6.22 is left as an exercise.

Exercises

6.4.1 Let $G = G(A, B, E)$ be a bipartite graph such that $|E| \geq |A||B|/K$. Show that there exists a subset A' of A of cardinality $|A'| \geq |A|/K$ such that any two elements in A' are connected by at least one path of length 2 in G. Show that $|A|/K$ cannot be improved to $|A|/K + 1$, even when A, B, and K are large.

6.4.2 [210] Let d be a large integer. Let $V = \{0, 1\}^d$, be the d-dimensional discrete cube, and let $G = G(V, E)$ be the bipartite graph formed by joining an edge between $x, y \in V$ if x and y differ in at most $d/2$ coordinates (i.e. if the Hamming distance between x and y is at most $d/2$). Show that $|E| = (\frac{1}{4} + o_{d\to\infty}(1))|V|^2$, but if V' is any subset of V with size $|V'| \geq c|V|$ then there exist x, x' in V' that are connected by fewer than $o_{d\to\infty}(|V|)$ paths of length 2 in G. (Hint: use a volume-packing argument to find two points x, x' in V' which are almost antipodal in the sense that their Hamming distance is $d - O(1)$.) Convert this example into a bipartite example and show that one cannot expect to eliminate the $(1 - \varepsilon)$ factor in Lemma 6.19 even if one lets ε be sufficiently small depending on K.

6.4.3 (Benny Sudakov, private communication) Let G be a bipartite graph $G = G(A, B, E)$ with $|A| = |B| = N$ and $|E| = \Theta(N^2)$ where N is sufficiently large. Show that G contains a complete bipartite graph with $\Omega(\log N)$ vertices in each color class. Show that the bound $\Omega(\log N)$ is best possible.

6.4.4 Let Z be the finite additive group $Z = \mathbf{Z}_2^d$ for some integer d, and let $\hat{Z}$ be the Pontryagin dual. Let $G = G(Z, \hat{Z}, E)$ be the bipartite graph formed by connecting $x \in Z$ to $\chi \in Z$ whenever $\chi(x) = 0$. Show that $|E| = |A||B|/2$. Using (4.2), show that one has $|A||B| \leq |Z|$ whenever $A \subseteq Z$, $B \subseteq Z'$ is a bipartite clique in G. Conversely, whenever N_1 and N_2 are positive integers such that $N_1 N_2 = |Z|$, show that there exists a bipartite clique $A \subseteq Z$, $B \subseteq Z'$ in G with $|A| = N_1$ and $|B| = N_2$. Compare this result with Exercise 6.4.3.

6.4.5 (Dyadic pigeonhole principle) Let $G = G(A, B, E)$ be a bipartite graph with $|E| \geq |A||B|/K$ for some $K \geq 1$. Show that there exists some $1 \leq K' \leq K$ and some induced subgraph $G' = G(A', B, E')$ of $G(A, B, E)$ with

$$|E|/(C + C\log K) \leq |E'| \leq |E|; \quad |A|/(C + C\log K) \leq |A'| \leq |A|$$

such that $|B|/2K' \leq \deg_{G'}(a) \leq |B|/K'$ for all $a \in A'$.

6.4.6 (Simultaneous popularity principle) Let $G = G(A, B, E)$ be a bipartite graph with $|E| \geq |A||B|/K$ for some $K \geq 1$. Show that there exists an induced subgraph $G' = G(A', B', E')$ with the bounds

$$|A'||B'| \geq |E'| \geq \frac{|A||B|}{2K^2}$$

$$|A'| \geq \frac{|A|}{K^2}$$

$$|B'| \geq \frac{|B|}{K^2}$$