paths of length two with both endpoints in V_1, including degenerate paths. Show, G also contains at least $|E|^4/|V_1||V_2|$ cycles of length four.

6.1.5 [198] Let $G(V_1, V_2, E)$ be a bipartite graph with V_1, V_2 non-empty. Show, for any $k \geq 1$, that G contains at least $|E|^{2k}/|V_1|^{k-1}|V_2|^k$ paths of length $2k$ with both endpoints in V_1, including degenerate paths, and also that G contains at least $|E|^{2k+1}/|V_1|^k|V_2|^k$ paths of length $2k+1$ from V_1 to V_2. (Hint: using the popularity principle, one can obtain lower bounds like this but losing an absolute constant depending on k. Then use the tensor power trick (as in Corollary 2.19) to remove this constant.)

6.1.6 Let $G = G(V, E)$ be a graph. Using the first moment method, show that G contains a bipartite subgraph $G'(A, B, E')$ with $|E'| \geq \frac{1}{2}|E|$. Give an example to show that the number $\frac{1}{2}$ cannot be replaced by any larger constant.

6.2 Independent sets, sum-free subsets, and Sidon sets

Intuitively one expects graphs with small degrees to have large independent sets. The following theorem, due to Turán, quantifies this intuition.

Theorem 6.1 (Turán's theorem) *Let $G = G(V, E)$ be a graph on n vertices. Then G contains an independent set of size at least $\sum_{v \in V} \frac{1}{\deg(v)+1}$. In particular, if G has maximal degree d, then G has an independent set of size at least $n/(d+1)$.*

Proof We shall use the probablistic method, or more precisely the first moment method. Let $\pi : V \to [1, n]$ be a bijection chosen uniformly at random. Let us call a vertex $v \in V$ *good* if it is larger than all its neighbors, in the sense that $\pi(w) < \pi(v)$ whenever $w \in N(v)$, and let S be the set of all good vertices. It is clear that S is an independent set. Also, for any $v \in V$, the probability that v is good can be easily verified to be $\frac{1}{\deg(v)+1}$. Thus by linearity of expectation (1.4) we have

$$\mathbf{E}(|S|) = \sum_{v \in V} \mathbf{P}(v \in S) = \sum_{v \in V} \frac{1}{\deg(v) + 1}$$

and so $|S| \geq \sum_{v \in V} \frac{1}{\deg(v)+1}$ with positive probability. The claim follows. □

6.2.1 Sum-free subsets

In 1965, Erdős and Moser [86] (see also [166], Problem C14) posed the following question. If $B \subset A$ are two additive sets, let us say that B is *sum-free* with respect to A if no element of A can be represented as the sum of two distinct elements of B. Given any additive set A, let $\phi(A)$ be the cardinality of the largest subset of A

which is sum-free with respect to A. Let $\phi(n)$ be the smallest value of $\phi(A)$ among all sets A of size n; thus $\phi(n)$ is the largest number such that every set A of n reals contains a subset of cardinality $\phi(n)$ which is sum-free with respect to A.

Note that it is important that we require the elements of B be distinct in order for this problem to be interesting. To see this, consider the set $A := 2^\wedge[1, n] = \{2, 2^2, \ldots, 2^n\}$. Clearly, if B is any subset of A of two or more elements, then there exists an element of A which is the sum of two (equal) elements in B.

It was remarked by Klarner (unpublished) and mentioned by Erdős in [86] that $\phi(n) = \Omega(\log n)$ for large n. The first published proof of this bound appeared in Choi's paper [55] about ten years later:

Theorem 6.2 *Let n be a large integer. Any set A of n real numbers contains a subset B of cardinality $\log n - O(1)$ which is sum-free with respect to A. In other words, $\phi(n) \geq \log n - O(1)$.*

Proof Let us first prove the claim for sets A of positive reals. Let us order the elements of A as $a_1 > a_2 > \cdots > a_n > 0$. Consider the graph G with vertices A, with two distinct elements $a, b \in A$ connected by an edge if and only if $a + b \in A$. By Theorem 6.1, this graph contains an independent vertex set B of size

$$|B| \geq \sum_{i=1}^{n} \frac{1}{\deg(a_i) + 1}.$$

Since B is independent in G, we see that B is sum-free with respect to A. Also, since $a_i + a_j > a_i$, and there are only $n - i$ elements of A larger than a_i, we see that $\deg(a_i) \leq n - i$ for all i. Since $\sum_{i=1}^{n} \frac{1}{n-i+1} = \log n - O(1)$, the claim follows.

To prove the general case, observe from the pigeonhole principle that any set of n reals either contains a subset of $n/2 - O(1)$ positive reals or $n/2 - O(1)$ negative reals, and the claim then follows (for large n) from the preceding paragraph. □

Let us now discuss the upper bound. Thus, we are interested in constructing sets A which do not contain large sum-free subsets. Erdős and Moser [86] proved that $\phi(n) \leq n/3$ and suggested that it probably has order $o(n)$. The first improvement over the Erdős and Moser result was due to Selfridge, who showed $\phi(n) \leq n/4$. Choi [55], using sieve methods, proved that $\phi(n) \leq O_\epsilon\left(n^{2/5+\epsilon}\right)$ for all $\epsilon > 0$. He also noted that in this problem it suffices to consider the special case when A is a set of positive integers. Choi's result was slightly improved by Baltz, Schoen and Srivastav [17], who showed that $\phi(n) \leq O(n^{2/5} \log^{2/5} n)$. A significant improvement of the upper bound was very recently obtained by Ruzsa [297] who proved that

$$\phi(n) = e^{O(\sqrt{\log n})}.$$

In the following we describe Ruzsa's construction, which, besides being very clever, is short and instructive. A key trick is to use a Freiman isomorphism to embed the problem in a very large-dimensional space (see also Exercise 10.1.4).

We shall need a dimension $d = \Theta(\sqrt{\log n})$. Using a Freiman isomorphism (see Lemma 5.25) it is enough to construct a set $A \subset \mathbf{Z}^d$ such that $|A| > n$ and $\phi(A) \leq e^{O(\sqrt{\log n})}$. For any $r > 0$, let $D_r \subset \mathbf{Z}^d$ be the set of integral lattice points in the ball of radius r centered at the origin, thus

$$D_r := \left\{ (x_1, \dots, x_d) \in \mathbf{Z}^d \mid \sum_{i=1}^{d} x_i^2 \leq r^2 \right\}.$$

We then set

$$A := \bigcup_{i=0}^{r-1} 2^i \cdot D_{r-i}$$

where $r = e^{O(\sqrt{\log n})}$. For an appropriate choice of d and r one can make $|A| > n$ and we claim that

$$\phi(A) \leq 2^d\, r = e^{O(\sqrt{\log n})}$$

Indeed, let $S \subset A$ have cardinality greater than $2^d r$. Then by the pigeonhole principle there exists $0 \leq i < r$ such that $|S \cap (2^i \cdot D_{r-i})| > 2^d$. Since $|D_1| = 2d + 1 < 2^d$, we see that $i < r - 1$. By the pigeonhole principle again, we can then find two vectors $s', s'' \in S \cap (2^i \cdot D_{r-i})$ which are congruent modulo $2 \cdot \mathbf{Z}^d$ (i.e. they have the same parity in each coordinate). Then one easily verifies that $s' + s'' \in 2^{i+1} \cdot D_{r-i-1} \subseteq A$, and so S is not sum-free with respect to A.

Remark 6.3 We return to the lower bound. In the same paper which established the upper bound, Ruzsa [297] improved Choi's result slightly by showing $\phi(n) > 2\log_3 n - 1$. Given the fact that Ruzsa's upper bound is sub-polynomial, one may suspect that $\phi(n) = \Theta(\log n)$, i.e., the right order of magnitude of $\phi(n)$ is $\log n$. It is, however, not the case. In a recent paper, Sudakov, Szemerédi and Vu [340] proved that $\phi(n)$ is super-logarithmic: thus in Landau notation

$$\phi(n) = \omega(n) \log n.$$

While this result improves Choi's result only slightly, its proof requires heavy machinery that involves the Balog–Szemerédi–Gowers theorem, Freiman's theorem, and Szemerédi's theorem. In this paper [340], the authors also proved a hypergraph version of the Balog–Szemerédi–Gowers theorem (see Section 6.4).