3 Alterations

We saw the alterations method in Section 1.1 to give lower bounds to Ramsey numbers. The basic idea is to first make a random construction, and then fix the blemishes.

3.1 Dominating set in graphs

In a graph $G = (V, E)$, we say that $U \subseteq V$ is ***dominating*** if every vertex in $V \setminus U$ has a neighbor in U.

Theorem 3.1.1

Every graph on n vertices with minimum degree $\delta > 1$ has a dominating set of size

$$\leq \left(\frac{\log(\delta+1)+1}{\delta+1}\right)n.$$

Naive attempt: take out vertices greedily. The first vertex eliminates $1+\delta$ vertices, but subsequent vertices eliminate possibly fewer vertices.

Proof. Two-step process (alteration method):
1. Choose a random subset
2. Add enough vertices to make it dominating

Let $p \in [0, 1]$ to be decided later. Let X be a random subset of V where every vertex is included with probability p independently.

Let $Y = V \setminus (X \cup N(X))$. Each $v \in V$ lies in Y with probability $\leq (1-p)^{1+\delta}$.

Then $X \cup Y$ is dominating, and

$$\mathbb{E}[|X \cup Y|] = \mathbb{E}[|X|] + \mathbb{E}[|Y|] \leq pn + (1-p)^{1+\delta}n \leq (p + e^{-p(1+\delta)})n$$

using $1 + x \leq e^x$ for all $x \in \mathbb{R}$. Finally, setting $p = \frac{\log(\delta+1)}{\delta+1}$ to minimize $p + e^{-p(1+\delta)}$, we bound the above expression by

$$\leq \left(\frac{1+\log(\delta+1)}{\delta+1}\right). \qquad \square$$

3.2 Heilbronn triangle problem

Question 3.2.1

How can one place n points in the unit square so that no three points forms a triangle with small area?

Let

$$\Delta(n) = \sup_{\substack{S \subseteq [0,1]^2 \\ |S|=n}} \min_{\substack{p,q,r \in S \\ \text{distinct}}} \text{area}(pqr)/$$

Naive constructions fair poorly. E.g., n points around a circle has a triangle of area $\Theta(1/n^3)$ (the triangle formed by three consectutive points has side lengths $\asymp 1/n$ and angle $\theta = (1 - 1/n)2\pi$). Even worse is arranging points on a grid, as you would get triangles of zero area.

Heilbronn conjectured that $\Delta(n) = O(n^{-2})$.

Komlós, Pintz, and Szemerédi (1982) disproved the conjecture, showing $\Delta(n) \gtrsim n^{-2} \log n$. They used an elaborate probabilistic construction. Here we show a much simpler version probabilistic construction that gives a weaker bound $\Delta(n) \gtrsim n^{-2}$.

***Remark* 3.2.2** (Upper bounds)**.** For a long time, the best upper bound known was $\Delta(n) \le n^{-8/7+o(1)}$ due to Komlós, Pintz, and Szemerédi (1981). This was recently improved to $\Delta(n) \le n^{-8/7-c}$ by Cohen, Pohoata, and Zakharov (2023+).

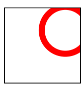
Theorem 3.2.3 (Many points without small area triangles)

For every positive integer n, there exists a set of n points in $[0,1]^2$ such that every triple spans a triangle of area $\ge cn^{-2}$, for some absolute constant $c > 0$.

Proof. Choose $2n$ points at random. For every three random points p, q, r, let us estimate

$$\mathbb{P}_{p,q,r}(\text{area}(p,q,r) \le \varepsilon).$$

By considering the area of a circular annulus around p, with inner and outer radii x and $x + \Delta x$, we find



$$\mathbb{P}_{p,q}(|pq| \in [x, x + \Delta x]) \le \pi((x + \Delta x)^2 - x^2).$$

So the probability density function satisfies

$$\mathbb{P}_{p,q}(|pq| \in [x, x+dx]) \le 2\pi x dx.$$

For fixed p, q

$$\mathbb{P}_r(\text{area}(pqr) \le \varepsilon) = \mathbb{P}_r\left(\text{dist}(pq, r) \le \frac{2\varepsilon}{|pq|}\right) \lesssim \frac{\varepsilon}{|pq|}.$$

Thus, with p, q, r at random

$$\mathbb{P}_{p,q,r}(\text{area}(pqr) \le \varepsilon) \lesssim \int_0^{\sqrt{2}} 2\pi x \frac{\varepsilon}{x}\, dx \asymp \varepsilon.$$

Given these $2n$ random points, let X be the number of triangles with area $\le \varepsilon$. Then $\mathbb{E}X = O(\varepsilon n^3)$.

Choose $\varepsilon = c/n^2$ with $c > 0$ small enough so that $\mathbb{E}X \le n$.

Delete a point from each triangle with area $\le \varepsilon$.

The expected number of remaining points is $\mathbb{E}[2n - X] \ge n$, and no triangles with area $\le \varepsilon = c/n^2$.

Thus with positive probability, we end up with $\ge n$ points and no triangle with area $\le c/n^2$. □

Algebraic construction. Here is another construction due to Erdős (in appendix of Roth (1951)) also giving $\Delta(n) \gtrsim n^{-2}$:

Let p be a prime. The set $\{(x, x^2) \in \mathbb{F}_p^2 : x \in \mathbb{F}_p\}$ has no 3 points collinear (a parabola meets every line in ≤ 2 points). Take the corresponding set of p points in $[p]^2 \subseteq \mathbb{Z}^2$. Then every triangle has area $\ge 1/2$ due to Pick's theorem. Scale back down to a unit square. (If n is not a prime, then use that there is a prime between n and $2n$.)

3.3 Markov's inequality

We note an important tool that will be used next.

Theorem 3.3.1 (Markov's inequality)
Let $X \ge 0$ be random variable. Then for every $a > 0$,

$$\mathbb{P}(X \ge a) \le \frac{\mathbb{E}[X]}{a}.$$

Proof. $\mathbb{E}[X] \geq \mathbb{E}[X 1_{X\geq a}] \geq \mathbb{E}[a 1_{X \geq a}] = a\mathbb{P}(X \geq a)$ □

Take-home message: for r.v. $X \geq 0$, if $\mathbb{E}X$ is *very* small, then *typically* X is small.

3.4 High girth and high chromatic number

If a graph has a k-clique, then you know that its chromatic number is at least k.

Conversely, if a graph has high chromatic number, is it always possible to certify this fact from some "local information"?

Surprisingly, the answer is no. The following ingenious construction shows that a graph can be "locally tree-like" while still having high chromatic number.

The ***girth*** of a graph is the length of its shortest cycle.

Theorem 3.4.1 (Erdős 1959)
For all k, ℓ, there exists a graph with girth $> \ell$ and chromatic number $> k$.

Proof. Let $G \sim G(n, p)$ with $p = (\log n)^2/n$ (the proof works whenever $\log n/n \ll p \ll n^{-1+1/\ell}$). Here $G(n, p)$ is Erdős–Rényi random graph (n vertices, every edge appearing with probability p independently).

Let X be the number of cycles of length at most ℓ in G. By linearity of expectations, as there are exactly $\binom{n}{i}(i-1)!/2$ cycles of length i in K_n for each $3 \leq i \leq n$, we have (recall that ℓ is a constant)

$$\mathbb{E}X = \sum_{i=3}^{\ell} \binom{n}{i} \frac{(i-1)!}{2} p^i \leq \sum_{i=3}^{\ell} n^i p^i = \ell (\log n)^{2i} = o(n).$$

By Markov's inequality

$$\mathbb{P}(X \geq n/2) \leq \frac{\mathbb{E}X}{n/2} = o(1).$$

(This allows us to get rid of all short cycles.)

How can we lower bound the chromatic number $\chi(\cdot)$? Note that $\chi(G) \geq |V(G)|/\alpha(G)$, where $\alpha(G)$ is the independence number (the size of the largest independent set).

With $x = (3/p)\log n = 3n/\log n$,

$$\mathbb{P}(\alpha(G) \geq x) \leq \binom{n}{x}(1-p)^{\binom{x}{2}} < n^x e^{-px(x-1)/2} = (ne^{-p(x-1)/2})^x = n^{-\Theta(n)} = o(1).$$

Let n be large enough so that $\mathbb{P}(X \geq n/2) < 1/2$ and $\mathbb{P}(\alpha(G) \geq x) < 1/2$. Then there is some G with fewer than $n/2$ cycles of length $\leq \ell$ and with $\alpha(G) \leq 3n/\log n$.

Remove a vertex from each cycle to get G'. Then $|V(G')| \geq n/2$, girth $> \ell$, and $\alpha(G') \leq \alpha(G) \leq 3n/\log n$, so

$$\chi(G') \geq \frac{|V(G')|}{\alpha(G')} \geq \frac{n/2}{3n/\log n} = \frac{\log n}{6} > k$$

if n is sufficiently large. $\square$

***Remark* 3.4.2.** Erdős (1962) also showed that in fact one needs to see at least a linear number of vertices to deduce high chromatic number: for all k, there exists $\varepsilon = \varepsilon_k$ such that for all sufficiently large n there exists an n-vertex graph with chromatic number $> k$ but every subgraph on $\lfloor \varepsilon n \rfloor$ vertices is 3-colorable. (In fact, one can take $G \sim G(n, C/n)$; see "Probabilistic Lens: Local coloring" in Alon–Spencer)

3.5 Random greedy coloring

In Section 1.3, we saw a simple argument showing that every k-uniform hypergraph with than 2^{k-1} edges is 2-colorable (meaning that we can color the vertices red/blue without no monochromatic edge). Take a moment to remember the proof.

In this section, we improve this result. The next result gives the current best known bound.

Theorem 3.5.1 (Radhakrishnan and Srinivasan (2000))

There is some constant $c > 0$ so that every k-uniform hypergraph with at most $c\sqrt{\frac{k}{\log k}}2^k$ edges is 2-colorable.

Recall from Section 1.3 that there exists a non-2-colorable k-uniform hypergraph on k^2 vertices and $O(k^2 2^k)$ edges, via a random construction.

Here we present a simpler proof, based on a ***random greedy coloring***, due to Cherkashin and Kozik (2015), following an approach of Pluhaár (2009).

Proof. Consider a k-graph with m edges.

Let us order the vertices using a uniformly random chosen permutation.

Color vertices greedily from left to right: color a vertex blue unless it would create a monochromatic edge, in which case color it red (i.e., every red vertex is the final vertex in an edge with all earlier $k-1$ vertices have already been colored blue).

The resulting coloring has no blue edges. The greedy coloring succeeds if it does not create a red edge.

Analyzing a greedy coloring is tricky, since the color of a single vertex may depend on the entire history. Instead, we identify a specific feature that necessarily results from a unsuccessful coloring.

If there is a red edge, then there must be two edges e, f so that the last vertex of e is the first vertex of f. Call such pair (e, f) ***conflicting*** (note that whether (e, f) is conflicting depends on the random ordering of the vertices, but not on how we assigned colors).

What is the probability of seeing a conflicting pair? Here is the randomness comes from the random ordering of vertices.

Each pair of edges with exactly one vertex in common conflicts with probability $\frac{(k-1)!^2}{(2k-1)!} = \frac{1}{2k-1}\binom{2k-2}{k-1}^{-1} \asymp k^{-1/2}2^{-2k}$. Summing over all $\leq m^2$ pairs of edges that share a unique vertex, we find that the expected number of conflicting pairs is at most $\lesssim m^2 k^{-1/2}2^{-2k}$, which is < 1 for some $m \asymp k^{1/4}2^k$. In this case, there is some ordering of vertices creating no conflicting pairs, in which case the greedy coloring always succeeds.

The above argument, due to Pluhaár (2009), yields $m \lesssim k^{1/4}2^k$. Next we will refine the argument to obtain a better bound of $\sqrt{\frac{k}{\log k}}2^k$ as claimed.

Instead of just considering a random permutation, let us map each vertex to $[0, 1]$ independently and uniformly at random. This map induces an ordering of the vertices, but it comes with further information that we will use.

Write $[0, 1] = L \cup M \cup R$ where (p to be decided)

$$L := \left[0, \frac{1-p}{2}\right), \qquad M := \left[\frac{1-p}{2}, \frac{1+p}{2}\right], \qquad R := \left(\frac{1+p}{2}, 1\right].$$

The probability that a given edge lands entirely in L is $(\frac{1-p}{2})^k$, and likewise with R. Taking a union bound over all edges,

$$\mathbb{P}(\text{some edge lies in } L \text{ or } R) \leq 2m\left(\frac{1-p}{2}\right)^k.$$

Suppose that no edge of H lies entirely in L or entirely in R. If (e, f) conflicts, then their unique common vertex $x_v \in e \cap f$ must lie in M. So the probability that (e, f)

conflicts is (here we use $x(1-x) \le 1/4$)

$$\int_{(1-p)/2}^{(1+p)/2} x^{k-1}(1-x)^{k-1}\,dx \le p4^{-k+1}.$$

Taking a union bound over all $\le m^2$ pairs of edges, we find that

$$\mathbb{P}(\text{some conflicting pair has the common vertex in } M) \le m^2 p 4^{-k+1}.$$

Thus

$$\begin{aligned}
&\mathbb{P}(\text{there is a conflicting pair}) \\
&\quad \le \mathbb{P}(\text{some edge lies in } L \text{ or } R) + \mathbb{P}(\text{some conflicting pair has the common vertex in } M) \\
&\quad \le 2m\left(\frac{1-p}{2}\right)^k + m^2 p 4^{-k+1} \\
&\quad < 2^{-k+1} m e^{-pk} + (2^{-k+1}m)^2 p
\end{aligned}$$

set $p = \log(2^{k-1}k/m)/k$ to minimize the right-hand side to get

$$= \frac{m^2}{4^{k-1}k} + \frac{m^2}{4^{k-1}k}\log\left(\frac{2^{k-2}k}{m}\right)$$

which is < 1 for $m = c2^k\sqrt{k/\log k}$ with $c > 0$ being a sufficiently small constant (we should assume that is k large enough to ensure $p \in [0,1]$; smaller values of k can be handled in the theorem exceptionally by later reducing the constant c). □

Food for thought: what is the source of the gain in the the $L \cup M \cup R$ argument? The expected number of conflicting pairs is unchanged. It must be that we are somehow clustering the bad events by considering the event when some edge lies in L or R.

It remains an intriguing open problem to improve this bound further.

Exercises

1. Using the alteration method, prove the Ramsey number bound

$$R(4,k) \ge c(k/\log k)^2$$

   for some constant $c > 0$.

2. Prove that every 3-uniform hypergraph with n vertices and $m \ge n$ edges contains an independent set (i.e., a set of vertices containing no edges) of size at least

$cn^{3/2}/\sqrt{m}$, where $c > 0$ is a constant.

3. Prove that every k-uniform hypergraph with n vertices and m edges has a transversal (i.e., a set of vertices intersecting every edge) of size at most $n(\log k)/k + m/k$.

4. *Zarankiewicz problem.* Prove that for every positive integers $n \geq k \geq 2$, there exists an $n \times n$ matrix with $\{0, 1\}$ entries, with at least $\frac{1}{2}n^{2-2/(k+1)}$ 1's, such that there is no $k \times k$ submatrix consisting of all 1's.

5. Fix k. Prove that there exists a constant $c_k > 1$ so that for every sufficiently large $n > n_0(k)$, there exists a collection $\mathcal{F}$ of at least c_k^n subsets of $[n]$ such that for every k distinct $F_1, \ldots, F_k \in \mathcal{F}$, all 2^k intersections $\bigcap_{i=1}^k G_i$ are nonempty, where each G_i is either F_i or $[n] \setminus F_i$.

6. *Acute sets in* $\mathbb{R}^n$. Prove that, for some constant $c > 0$, for every n, there exists a family of at least $c(2/\sqrt{3})^n$ subsets of $[n]$ containing no three distinct members A, B, C satisfying $A \cap B \subseteq C \subseteq A \cup B$.
   Deduce that there exists a set of at least $c(2/\sqrt{3})^n$ points in $\mathbb{R}^n$ so that all angles determined by three points from the set are acute.
   Remark. The current best lower and upper bounds for the maximum size of an "acute set" in $\mathbb{R}^n$ (i.e., spanning only acute angles) are $2^{n-1} + 1$ and $2^n - 1$ respectively.

7. $\star$ *Covering complements of sparse graphs by cliques*

   a) Prove that every graph with n vertices and minimum degree $n - d$ can be written as a union of $O(d^2 \log n)$ cliques.

   b) Prove that every bipartite graph with n vertices on each side of the vertex bipartition and minimum degree $n - d$ can be written as a union of $O(d \log n)$ complete bipartite graphs (assume $d \geq 1$).

8. $\star$ Let $G = (V, E)$ be a graph with n vertices and minimum degree $\delta \geq 2$. Prove that there exists $A \subseteq V$ with $|A| = O(n(\log \delta)/\delta)$ so that every vertex in $V \setminus A$ contains at least one neighbor in A and at least one neighbor not in A.

9. $\star$ Prove that every graph G without isolated vertices has an induced subgraph H on at least $\alpha(G)/2$ vertices such that all vertices of H have odd degree. Here $\alpha(G)$ is the size of the largest independent set in G.