3 Alterations

Recall the naive probabilistic method: we found some lower bounds for Ramsey numbers in Section 1.1, primarily for the diagonal numbers. We did this with a basic method: color randomly, so that we color each edge red with probability p and blue with probability $1-p$. Then the probability that we don't see any red s-cliques or blue t-cliques (with a union bound) is at most

$$\binom{n}{s}p^{\binom{s}{2}} + \binom{n}{t}(1-p)^{\binom{t}{2}},$$

and if this is less than 1 for some p, then there exists some graph on n vertices for which there is no red K_s and blue K_t. So we union bounded the bad events there.

Well, the alteration method does a little more than that - here's a proof that mirrors that of Proposition 1.6. We again color randomly, but the idea now is to delete a vertex in every bad clique (red K_s and blue K_t). How many edges have we deleted? We can estimate by using linearity of expectation:

Theorem 3.1

For all $p \in (0,1)$, $n \in \mathbb{N}$,

$$R(s,t) > n - \binom{n}{s}p^{\binom{s}{2}} - \binom{n}{t}(1-p)^{\binom{t}{2}}.$$

This right hand side begins by taking the starting number of vertices and then we deleting one vertex for each clique. We're going to explore this idea of "fixing the blemishes" a little more.

3.1 Dominating sets

Definition 3.2

Given a graph G, a **dominating set** U is a set of vertices such that every vertex not in U has a neighbor in U.

Basically, we want a subset of vertices such that every vertex is either picked or adjacent to something we picked. Clearly the whole set of vertices is dominating, but our goal is to find small dominating sets relative to the number of vertices.

Theorem 3.3

If our graph G has n vertices and minimum degree δ among all vertices ($\delta > 1$), then G has a dominating set of size at most $\frac{\log(\delta+1)+1}{\delta+1}n$.

Proof. We will do a two-step process. First, pick a random subset X by including every vertex with probability p. Then, add all vertices that are neither in X or the neighbors of X (since those are the ones we haven't covered with our set yet); call this Y. By this point, we have a dominating set $X \cup Y$ by construction.

Now, how many vertices do we have in our dominating set? Any vertex v is in Y if neither v nor any of its neighbors are in X. So v has probability $(1-p)^{\deg(v)+1} \le (1-p)^{1+\delta}$ of being included in Y, meaning that the expected size of $X \cup Y$ is

$$\mathbb{E}[X] + \mathbb{E}[Y] = np + n(1-p)^{1+\delta}.$$

Now we just optimize for p. The important computational trick is that we can bound this pretty well if p is small:

$$\leq np + ne^{-p(1+\delta)}.$$

Turns out the optimal value is $p = \frac{\log(\delta+1)}{\delta+1}$, and this gives the result we want. $\square$

3.2 A problem from discrete geometry

Problem 3.4 (Heilbronn triangle problem)

Place n points in the unit square. How large can we make the smallest area of any triangle formed by our points?

This is related to the ideas of **discrepancy theory**. There are applications when we want to evenly distribute points, and this is one way of quantifying that randomness.

Definition 3.5

Let $\Delta(n)$ be the minimum real number such that for any n points in the unit square, there are three points with triangle area at most $\Delta(n)$.

For example, it's bad to have a square grid of points, since we get a minimal area of 0. If we put the n points on a circle, we get an area on the order of $\frac{1}{n^3}$, which is at least nonzero. The whole point is that we don't want collinearity, so it's hard to think about an efficient picture that is "irregular."

Heilbronn conjectured that $\Delta(n) \lesssim n^{-2}$, but this was disproved in 1982 by KPS: they showed $\Delta(n) \gtrsim \frac{\log n}{n^2}$. On the other hand, the best known upper bound is $\lesssim n^{-\frac{8}{7}+o(1)}$.

Below, we use a randomized construction to show that $\Delta(n) \gtrsim n^{-2}$:

Proposition 3.6

There exist n points in a unit square such that every three form a triangle with area at least cn^{-2} for some constant $c > 0$.

Proof. Choose $2n$ points at random (uniformly in the unit square). How can we find the probability that the area of a triangle pqr is at most ε?

Pick p first. The probability that the distance between p and q is in the range $[x, x+\Delta x]$ is the intersection of the square and the annulus with bounds x and $x+\Delta x$, which is always at most $\Theta(x\Delta x)$ (by taking Δx to be small).

So now, if we fix p and q, what's the probability that our area is less than ε; that is, the height from r to line pq is small? This means we want the distance between line pq and point r to be at most $\frac{2\varepsilon}{\text{dist}(p,q)}$, which is bounded by a constant times $\frac{\varepsilon}{x}$ (because the allowed region is bounded by a rectangle with height $\frac{4\varepsilon}{x}$ and length $\sqrt{2}$).

Putting these together, the probability that the area is at most ε can be bounded by a factor proportional to

$$\int_0^{\sqrt{2}} x \cdot \frac{\varepsilon}{x} dx \lesssim \varepsilon.$$

So now we apply the idea of the alteration method: let X be the number of triangles with area ε, and delete 1 point from each triangle: let's say we delete x triangles. What's the expected number of points that are removed? We remove $\mathbb{E}[X] \propto \varepsilon n^3$ points, and we'll pick $\varepsilon = \frac{c}{n^2}$ for some constant c such that the expected value of x is $\leq n$. Now with positive probability, our process deleted fewer than n points, so we have at least n points with no small triangles of area less than $\frac{c}{n^2}$, and we're done. $\square$

Actually, we can also do a direct algebraic construction. Let's say we want to find n points in a square grid with no three points collinear. Note that a lattice polygon has area at least $\frac{1}{2}$, so take $n = p$ to be a prime number, and let our points be $\{(x, x^2) : x \in \mathbb{F}_{p^2}\}$ in $\mathbb{F}_p^2$. Parabolas have no three points collinear, and thus we've constructed configurations with smallest area proportional to n^{-2} explicitly.

So the idea is that although algebra solutions are pretty, it's often hard to modify algebraic constructions, while combinatorial proofs let us use heavier hammers.

3.3 Hard-to-color graphs

There are many problems in combinatorics for which probabilistic constructions are the only ones we know. Here's an example that Erdős studied.

Definition 3.7

The **chromatic number** $\chi(G)$ of a graph is the minimum number of colors needed to properly color G.

If we look at a very large graph and look at it locally, we can set some lower bounds on the chromatic number. For example, a K_4 means that $\chi(G) \geq 4$. Our question: is it possible to use local information to find that $\chi(G)$ is upper-bounded? Turns out the answer is no!

Definition 3.8

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

Theorem 3.9 (Erdős)

For all positive integers k and ℓ, there exists a graph of girth more than ℓ and chromatic number more than k.

The idea is that for graphs with large girth, we only see trees locally, and that won't tell us anything. So the chromatic number is (in some sense) a global statistic!

Theorem 3.10 (Markov's inequality)

Given a random variable X that only takes on nonnegative values, for all $a > 0$,

$$\Pr(X \geq a) \leq \frac{\mathbb{E}[X]}{a}.$$

Proof.

$$\mathbb{E}[X] \geq \mathbb{E}\left[X \cdot 1_{X \geq a}\right] \geq \mathbb{E}\left[a 1_{X \geq a}\right] = a \Pr(X \geq a).$$

□

This is used with the mindset that if the expected value of X is small, then X is small with high probability.

Proof of Theorem 3.9. Construct an **Erdős-Renyi random graph** $G(n, p)$ with n vertices and each edge appearing with probability p. Here, let's let

$$p = n^{\theta - 1}, 0 < \theta < \frac{1}{\ell}.$$

Let X be the number of cycles of length at most ℓ. By expected value calculations, the number of such cycles is

$$\mathbb{E}[X] = \sum_{i=3}^{\ell} \binom{n}{i} \frac{(i-1)!}{2} p^i$$

since given any i vertices, there are $\frac{(i-1)!}{2}$ different cycles through them. This can be upper bounded by

$$\leq \sum_{i=3}^{\ell} n^i p^i \leq \ell n^{\ell} p^{\ell}.$$

Plugging in our choice of p, this evaluates to

$$\ell n^{\theta \ell} = o(n)$$

by our choice of θ. Now, what's the probability we have lots of short cycles? By Markov's inequality,

$$\Pr\left(X \geq \frac{n}{2}\right) \leq \frac{\mathbb{E}[X]}{n/2} = o(1),$$

so this allows us to find a graph with no cycles of length at most ℓ by the alteration method.

Meanwhile, what about the chromatic number? The easiest way to lower bound the chromatic number is to upper bound the independence number $\alpha(G)$, which is the size of the largest independent set. Note that every color class is an independent set (since no two vertices with the same color share an edge), so

$$|V(G)| \leq \chi(G)\alpha(G),$$

which is good for us as it gives a lower bound on the chromatic number. Well, the probability that we can have an independent set of size at least x is

$$\Pr(\alpha(G) \geq x) \leq \binom{n}{x}(1-p)^{\binom{x}{2}},$$

and if this quantity is small, we're good to lower bound the chromatic number. With more bounding,

$$\Pr(\alpha(G) \geq x) < n^x e^{-px(x-1)/2} = (ne^{-p(x-1)/2})^x$$

and by setting $x = \frac{3}{p}\log n$, this quantity becomes $o(1)$ as well.

We're almost done. Let n be large enough so that we have few cycles and large independent set size with high probability: $X \leq \frac{n}{2}$ and $\alpha \geq x$, each with probability greater than $\frac{1}{2}$. There now exists G with at least $\frac{n}{2}$ cycles of length ℓ and $\alpha(G) \leq \frac{3}{p}\log n$, and now remove a vertex from each short cycle (of length ℓ) to get a graph G'. The number of vertices of G' is now at least $\frac{n}{2}$, since we only removed at most $\frac{n}{2}$ cycles worth of vertices, and

$$\alpha(G') \leq \alpha(G) \leq \frac{3}{p}\log n,$$

so

$$\chi(G') \geq \frac{|V(G')|}{\alpha(G')} \geq \frac{np}{6\log n} = \frac{n^{\theta}}{6\log n} > k$$

for some sufficiently large n, and therefore G' is the graph we're looking for. □

3.4 Coloring edges

Recall that we defined $m(k)$ in Section 1.5 to be the minimum number of edges in a k-uniform hypergraph that is not 2-colorable. (Basically, we want to color the vertex sets red and blue so that no edge is monochromatic.) We found an upper and lower bound earlier: a randomized construction gives $m(k) \lesssim k^2 2^k$ using k^2 vertices, and $m(k) \geq 2^{k-1}$, just by randomly coloring the vertices, since each edge fails with some probability. Let's improve this lower bound now:

Theorem 3.11

$$m(k) \gtrsim \sqrt{\frac{k}{\log k}} 2^k.$$

Proof. Let's say a hypergraph H has m edges. Consider a random greedy coloring: choose a random mapping of the vertices to $[0, 1]$, and go from left to right, always coloring blue unless we would create a blue edge (in which case we color red).

What's the probability this gives a proper coloring? The only possible failures are red edges: call two edges e and f **conflicting** if they share exactly one vertex, and that vertex is the final vertex of e and first vertex of f. The idea here is that any failure must give a pair of conflicting edges.

So what's the probability that such a pair exists? Let's bound it: given two edges e and f that share exactly one vertex, the probability that they conflict is

$$P(e, f) = \frac{(k-1)!^2}{(2k-1)!} = \frac{1}{(2k-1)\binom{2k-2}{k-1}}.$$

Asymptotically, $\binom{n}{n/2}$ is $\frac{2^n}{\sqrt{n}}$ up to a constant factor, so the probability that these two edges conflict is $\Theta\left(\frac{1}{2^{2k}\sqrt{k}}\right)$. Now if $P(e, f)$ is less than $\frac{1}{m^2}$, we're happy, because there's less than m^2 edges and we can union bound the bad events. Doing some algebra, this gives

$$m(k) \gtrsim k^{1/4} 2^k.$$

Now let's be more clever. Split the interval $[0, 1]$ into $L = \left[0, \frac{1-p}{2}\right]$, $M = \left[\frac{1-p}{2}, \frac{1+p}{2}\right]$, $R = \left[\frac{1+p}{2}, 1\right]$. A pair of edges that conflict must have $e \subseteq L$, $e \subseteq R$, $f \subseteq L$, or $f \subseteq R$, or they both intersect in the middle.

The probability that e lies in L is just $\left(\frac{1-p}{2}\right)^k$ (each of the k vertices must be in L), and we can say similar things about the cases $e \subseteq R, f \subseteq L, f \subseteq R$. To deal with the middle intersection, if the common vertex between e and f is v, the probability that the second scenario happens is the probability that there are $(k-1)$ vertices to the left of v in M for e and $(k-1)$ vertices to the right of v in M for f. This is bounded by

$$\int_{(1-p)/2}^{(1+p)/2} x^k (1-x)^{k-1} dx \leq p \left(\frac{1}{4}\right)^{k-1}.$$

Putting all of this together, the probability of any pair of conflicting edges is bounded by

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

and this is less than 1 if $m = c2^k \sqrt{\frac{k}{\log k}}$ and $p = \left(\log \frac{4m}{2^k}\right)/k$, and we've found a bound on m as desired. ☐