Random Graphs — Week 6

Yuval Filmus

November 30, 2019

1 Ramsey's theorem

A well-known puzzle asks to show that among six people, there are either three mutual friends or three mutual enemies. Ramsey's theorem is a far-reaching generalization. The Ramsey number $R(a, b)$ is the minimal n such that every graph on n vertices either contains a clique of size a or an independent set of size b.

Theorem 1.

$$R(\aleph_0, \aleph_0) = \aleph_0.$$

Proof. The lower bound is obvious. Let v_0 be an arbitrary vertex. Out of the infinitely many remaining vertices, either there are infinitely many neighbors of v_0, or infinitely many non-neighbors of v_0. In the former case, we color v_0 blue, and remove all non-neighbors. In the latter case, we color v_0 red, and remove all neighbors. Now let v_1 be an arbitrary vertex different from v_0, and repeat the same process.

In this way, we construct an infinite sequence $v_0, v_1, v_2, \ldots$ of vertices. There are either infinitely many blue vertices or infinitely many red vertices. By construction, any two blue vertices are connected to one another, and any two red vertices are not connected to one another. Hence we either get an infinite clique, or an infinite independent set. □

The infinite version of Ramsey's theorem implies, via compactness, that $R(a, b) < \infty$ for any $a, b < \infty$. Following essentially the same argument as the proof of the infinite Ramsey theorem, we can get concrete bounds.

Theorem 2.

$$R(a, b) \leq \binom{a+b-2}{a-1}.$$

Proof. The proof is by induction on $\min(a, b)$. If $a = 1$ then clearly we can take $n = 1$. Suppose now that $a, b \geq 2$, and consider a graph with $\binom{a+b-2}{a-1}$ vertices. Fix some vertex v. Since

$$\binom{a+b-2}{a-1} = \binom{a+b-3}{a-2} + \binom{a+b-3}{a-1},$$

either v has at least $\binom{a+b-3}{a-2}$ neighbors, or it has at least $\binom{a+b-3}{a-1}$ non-neighbors.

In the first case, consider the graph induced by the neighbors of v. By induction, it contains either a clique of size $a - 1$ or an independent set of size b. In the former case, we can add v to obtain a clique of size a, and in the latter case, we already have an independent set of size b.

The second case is similar. Considering the graph induced by the non-neighbors of v, we obtain either a clique of size a or an independent set of size $b-1$, which together with v forms an independent set of size b. $\square$

Taking $a = b = 3$, we obtain $R(3,3) \le \binom{4}{2} = 6$, which is the statement of the puzzle. This is tight: C_5 is a graph on five vertices containing neither a triangle nor a triangle of missing edges (to see this, note that the complement of C_5 is just another C_5).

The most interesting case of Ramsey's theorem is when $a = b$. In this case, the upper bound in the theorem is

$$\binom{2a-2}{a-1} = \Theta\left(\frac{4^a}{\sqrt{a}}\right).$$

Turning the theorem around, it shows that every graph on n vertices contains either a clique or an independent set of size roughly $\frac{1}{2}\log n$ (logarithm is base 2). Is this tight?

We can use a random graph to show that the bound is tight up to constant factors. The expected number of k-cliques in $G(n, 1/2)$ is

$$\binom{n}{k} 2^{-\binom{k}{2}} \le \left(\frac{en}{k}\right)^k 2^{-k(k-1)/2} = \left(\frac{e\sqrt{2}n}{k 2^{k/2}}\right)^k.$$

If $k = 2\log n$ then this quantity is $o(1)$, and so with high probability $G(n, 1/2)$ doesn't contain a clique of size $2\log n$. Due to symmetry, the same holds for independent sets. In particular, there exists a graph on n vertices with neither a clique nor an independent set of size $2\log n$.

Amazingly, no explicit construction of such a graph is known! However, it is conjectured that the *Paley* graph is a Ramsey graph (its clique number and independence number are both $O(\log n)$). For an odd prime p, the Paley graph has vertices $1, \ldots, p-1$, and we connect two vertices i, j if their difference $i - j$ is a quadratic residue modulo p, that is, if there exists k such that $k^2 \equiv i - j \pmod{p}$. It is not hard to check that exactly half the integers in $1, \ldots, p-1$ are quadratic residues, and so the density of the Paley graph is $1/2$. The Paley graph is known to be *quasirandom*, that it, for every graph H, the number of copies of H is close to the expected number of copies in $G(n, 1/2)$.

The best constructions of graphs with no large cliques or independent sets employ *non-malleable extractors*. The record-holder is Xin Li, who constructed a graph in which there are no cliques or independent sets of size $(\log n)^{C \log^{(3)} n / \log^{(4)} n}$, where $\log^{(k)}$ is the logarithm function, iterated k times.

2 Clique number

Let N_k be the expected number of k-cliques in $G(n, 1/2)$:

$$N_k = \binom{n}{k} 2^{-\binom{k}{2}} = \frac{n^{\underline{k}}}{k!} 2^{-k\frac{k-1}{2}}.$$

We will be interested for values of k around $2\log n$. For such values,

$$\frac{n^{\underline{k}}}{n^k} = \frac{n}{n} \cdot \frac{n-1}{n} \cdots \frac{n-k+1}{n} = \left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right) \geq 1 - \frac{k(k-1)/2}{n} = 1 - o(1).$$

Applying Stirling's approximation, we obtain

$$N_k \sim \frac{n^k}{\sqrt{2\pi k}(k/e)^k} = \frac{1}{\sqrt{2\pi k}} \left(\frac{e\sqrt{2}n}{k2^{k/2}}\right)^k.$$

We can guess that the critical value of k is obtained around the solution to $k2^{k/2} = n$, which is $k \approx 2\log n - 2\log\log n$. Indeed, if $k = 2\log n - 2\log\log n + C$ then

$$k2^{k/2} \sim 2\log n \cdot \frac{n}{\log n} \cdot 2^C = 2^{C+1}n,$$

and so

$$N_k \sim \frac{((e/\sqrt{2})/2^C)^k}{\sqrt{2\pi k}}.$$

This shows that the maximal k such that $N_k \geq 1$ is

$$k_0 = 2\log n - 2\log\log n + O(1).$$

Indeed, if $C = 1$ then $(e/\sqrt{2})/2^C < 1$, and so $N_k = o(1)$, while if $C = 0$ then $(e/\sqrt{2})/2^C > 1.9$ and so $N_k \gtrsim 1.9^k/\sqrt{2\pi k} \to \infty$.

As we have seen above, if k is within constant distance of k_0 then $k2^{k/2} = \Theta(n)$, and so

$$\frac{N_{k+1}}{N_k} = \frac{n-k}{k+1}2^{-k} \approx \frac{n}{k2^k} = \frac{nk}{(k2^{k/2})^2} = \Theta\left(\frac{nk}{n^2}\right) = \Theta\left(\frac{\log n}{n}\right).$$

In other words, the expected number of k-cliques drops sharply as we increase k. In particular, this implies that

$$N_{k_0-1} = \Omega\left(\frac{n}{\log n}\right), \quad N_{k_0+2} = O\left(\frac{\log n}{n}\right).$$

Thus, with high probability, $G(n, 1/2)$ doesn't contain a $(k_0 + 2)$-clique. In contrast, a second moment argument shows that with high probability, $G(n, 1/2)$ does contain a $(k_0 - 1)$-clique, an argument which we outline below. This shows that with high probability, $G(n, 1/2)$ does contain a (k_0-1)-clique. Since with high probability it doesn't contain a $(k_0 + 2)$-clique, we conclude that the clique number is, with high probability, one of $k_0 - 1, k_0, k_0 + 1$. With slightly more finesse, one is able to whittle down this list to only two values, and for most n, to only one value.

If X_k is the number of k-cliques, then enumerating the size r of the intersection of two k-cliques, we get

$$\mathbb{E}[X_k^2] = \sum_{r=0}^{k} \binom{n}{k}\binom{k}{r}\binom{n-k}{k-r}(1/2)^{2\binom{k}{2}-\binom{r}{2}}.$$

We are interested in

$$\frac{\mathbb{E}[X_k^2]}{\mathbb{E}[X_k]^2} = \sum_{r=0}^{k} \frac{\binom{k}{r}\binom{n-k}{k-r}}{\binom{n}{k}} 2^{\binom{r}{2}},$$

which we want to show is $1 + o(1)$. Note that the coefficients $\binom{k}{r}\binom{n-k}{k-r}/\binom{n}{k}$ are those of a hypergeometric distribution, corresponding to the following process: given that we have identified k vertices as one k-clique, we are again drawing k vertices, and r is the number of those which belong to the first clique.

Let us denote the r'th summand by J_r. We have

$$\begin{aligned}
J_0 &= \frac{\binom{n-k}{k}}{\binom{n}{k}} \approx \left(1 - \frac{k}{n}\right)^k \approx 1 - \frac{k^2}{n} = 1 - O\left(\frac{\log^2 n}{n}\right), \\
J_1 &= \frac{k\binom{n-k}{k-1}}{\binom{n}{k}} = \frac{k\binom{n-k}{k-1}}{\frac{n}{k}\binom{n-1}{k-1}} \approx \frac{k^2}{n}\left(1 - \frac{k}{n}\right)^{k-1} = O\left(\frac{\log^2 n}{n}\right), \\
J_k &= \frac{2^{\binom{k}{2}}}{\binom{n}{k}} = \frac{1}{N_k} = O\left(\frac{\log n}{n}\right),
\end{aligned}$$

the last estimate holding for $k = k_0 - 1$.

In the homework assignment, you will show that the sequence $J_0, \ldots, J_k$ is unimodal, implying that $\mathbb{E}[X_k^2]/\mathbb{E}[X_k]^2 = 1 + o(1)$, as needed.