5 High Girth and High Chromatic Number

Many consider the following one of the most pleasing uses of the probabilistic method, as the result is surprising and does not appear to call for nonconstructive techniques. The *girth* of a graph G is the size of its smallest circuit.
Theorem 5.1(Erdös [1959]). For all k, l there exists a graph G with $girth(G) > l$ and $\chi(G) > k$.
Proof. Fix $\theta < 1/l$ and let $G \sim G(n,p)$ with $p = n^{\theta-1}$. Let X be the number of circuits of size at most l. Then

$$E[X] = \sum_{i=3}^{l} \frac{(n)_i}{2i} p^i \leq \sum_{i=3}^{l} \frac{n^{\theta i}}{2i} = o(n)$$

as $\theta l < 1$. In particular

$$\Pr[X \geq \frac{n}{2}] = o(1)$$

Set $x = \lceil \frac{3}{p} \ln n \rceil$ so that

$$\Pr[\alpha(G) \geq x] \leq \binom{n}{x} (1-p)^{\binom{x}{2}} < \left[ne^{-p(x-1)/2}\right]^x = o(1)$$

Let n be sufficiently large so that both these events have probability less than .5. Then there is a specific G with less than $n/2$ cycles of length less than l and with $\alpha(G) < 3n^{1-\theta} \ln n$. Remove from G a vertex from each cycle of length at most l. This gives a graph G^* with at least $n/2$ vertices. G^* has girth greater than l and $\alpha(G^*) \leq \alpha(G)$. Thus

$$\chi(G^*) \geq \frac{|G^*|}{\alpha(G^*)} \geq \frac{n/2}{3n^{1-\theta} \ln n} = \frac{n^\theta}{6 \ln n}$$

To complete the proof, let n be sufficiently large so that this is greater than k.

Ramsey function. To unravel the definition $R(k,l) > n$ means that there *exists* a Red-Blue coloring of K_n with neither Red K_k nor Blue K_l. In his 1947 paper Erdős considered the case $k = l$.

Theorem 4.1. If

$$\binom{n}{k} 2^{1-\binom{k}{2}} < 1$$

then $R(k,l) > n$.

Proof. Let $G \sim G(n, \frac{1}{2})$ and consider the random two-coloring given by coloring the edges of G red and the other edges of K_n blue. Let X be the number of monochromatic K_k. Then the left hand side above is simply $E[X]$. With $E[X] < 1$, $\Pr[X = 0] > 0$. Hence there is a point in the probability space - i.e., a graph G, whose coloring has $X = 0$ monochromatic K_k. □

Note here a subtle (for some) point. With positive probability $G(n, \frac{1}{2})$ has the desired property and therefore there must - absolutely, positively - exist a G with the desired property. Random Graphs and the Probabilistic Method are closely related. In Random Graphs we study the probability of $G(n,p)$ having certain properties. In the Probabilistic Method our goal is to prove the existence of a G having certain properties. We create a probability space in which the probability of the random G having these properties is positive, and from that it follows that some such G must exist.

Applying some simple asymptotics to the theorem yields that $R(k,k) > \sqrt{2}^{n(1+o(1))}$. In 1935 Erdős and George Szekeres found the upper bound $R(k,k) < 4^{n(1+o(1))}$ by nonrandom means. While there have been improvements in lower order terms, these bounds remain the best known up to $(1+o(1))^n$ terms. It is also interesting that no exponential lower bound is known by constructive means.

A general lower bound is the following.

Theorem 4.2. If there exists $p \in [0,1]$ with

$$\binom{n}{k} p^{\binom{k}{2}} + \binom{n}{l} (1-p)^{\binom{l}{2}} < 1$$

then $R(k,l) > n$.

Proof. Let $G \sim G(n,p)$ and color the edges of G red and the other edges of K_n blue. Then the left hand side above is simply the expectation of the number of red K_k plus the number of blue K_l. For some G this is zero and that G gives the desired coloring. □

Dealing with the asymptotics of this result can be quite tricky. For example, what does this imply about $R(k,2k)$?

expectation $E[X] = \mu$. Now consider the r-th factorial moment $E[(X)_r]$ for any fixed r. By the symmetry $E[(X)_r] = (n)_r E[X_1 \cdots X_r]$. For vertices $1, \ldots, r$ to all be isolated the $r(n-1) - \binom{r}{2}$ pairs $\{i, x\}$ overlapping $1, \ldots, r$ must all not be edges. Thus

$$E[(X)_r] = (n)_r (1-p)^{r(n-1)-\binom{r}{2}} \sim n^r (1-p)^{r(n-1)} \sim \mu^r$$

(That is, the dependence among the X_i was asymptotically negligible.) As all the moments of X approach those of $P(\mu)$, X approaches $P(\mu)$ in distribution and in particular the theorem holds. □

Now we give the Erdős-Rényi famous "double exponential" result.

Theorem 3.2. Let

$$p = p(n) = \frac{\log n}{n} + \frac{c}{n} + o(\frac{1}{n})$$

Then

$$\lim_{n \to \infty} \Pr[G(n,p) \text{ is connected}] = e^{-e^{-c}}$$

For such p, $n(1-p)^{n-1} \sim \mu = e^{-c}$ and by the above argument the probability that X has no isolated vertices approaches $e^{-\mu}$. If G has no isolated vertices but is not connected there is a component of k vertices for some $2 \leq k \leq \frac{n}{2}$. Letting B be this event

$$\Pr[B] \leq \sum_{k=2}^{n/2} \binom{n}{k} k^{k-2} p^{k-1} (1-p)^{k(n-1)-\binom{k}{2}}$$

The first factor is the choice of a component set $S \subset V(G)$. The second factor is a choice of tree on S. The third factor is the probability that those tree pairs are in $E(G)$. The final factor is that there be no edge from S to $V(G) - S$. Some calculation (which we omit but note that $k = 2$ provides the main term) gives that $\Pr[B] = o(1)$ so that $X \neq 0$ and connectivity have the same limiting probability. □

4 The Probabilistic Method

In 1947 Paul Erdős started what is now called the Probabilistic Method with a three page paper in the Bulletin of the American Mathematical Society. The Ramsey function $R(k,l)$ is defined as the least n such that if the edges of K_n are colored Red and Blue then there is either a Red K_k or a Blue K_l. The existence of such an n is a consequence of Ramsey's Theorem and will not concern us here. Rather, we are interested in lower bounds on the

$\omega(\overline{G}) \leq (2\log_2 n)(1+o(1)$. This will turn out to be the right asymptotic answer.

For the lower bound (which is not best possible) we outline an analysis of the following "greedy algorithm". We find an independent set C on G as follows. Set $S_0 = V(G)$, $a_1 = 1$ and S_1 equal the set of vertices not adjacent to a_1. Having determined $a_1, \ldots, a_i$ and S_i let a_{i+1} be the least vertex of S_i and let S_{i+1} be those $x \in S_i - \{a_i\}$ not adjacent to a_{i+1}. Continue until $S_t = \emptyset$ and set $C = \{a_1, \ldots a_t\}$. A fairly straightforward analysis gives that $|C| \sim \log_2 n$ almost surely, and moreover that the probability (for any given $\epsilon > 0$ that $|C| < (log_2 n)(1-\epsilon)$ is $o(n^{-1})$. Call this one pass of the algorithm. Now we give all points of C color "one", remove vertices C from G and iterate. Let G^1 be G with C removed. Critically, it finding C we only "exposed" edges involving C so that we can consider G^1 to have distribution $G(n_1, \frac{1}{2})$, where $n_1 = n - |C|$ is the number of vertices. Letting n_j be the number of vertices remaining after the j-th pass, almost surely we have $n_{j+1} < n_j - (1-\epsilon)\log_2 n_j$ so that the algorithm is completed using less than $\frac{n}{\log_2 n}(1+\epsilon')$ colors. (Actually, to avoid end effects we can stop the algorithm when there are $o(n/\log n)$ vertices remaining and simply give each such vertex a separate color.)

It is tempting to improve the lower bound as follows. We know that almost surely G contains an independent set of size $\sim 2\log_2 n$. Let C be that set, remove C from G giving G^1 and iterate. The problem is, of course, that G^1 no longer has distribution $G(n_1, \frac{1}{2})$ and no proof has been found along these lines of the true result that $\chi(G) \sim \frac{n}{2\log_2 n}$ almost surely.

3 Connectivity

In this section we give a relatively simple example of what we call the Poisson Paradigm: the rough notion that if there are many rare and nearly independent events then the number of events that hold has approximately a Poisson distribution. This will yield one of the most beautiful of the Erdős-Rényi results, a quite precise description of the threshold behavior for connectivity. A vertex $v \in G$ is *isolated* if it is adjacent to no $w \in V$. In $G(n,p)$ let X be the number of isolated vertices.

Theorem 3.1. Let $p = p(n)$ satisfy $n(1-p)^{n-1} = \mu$. Then

$$\lim_{n\to\infty} \Pr[X = 0] = e^{-\mu}$$

We let X_i be the indicator random variable for vertex i being isolated so that $X = X_1 + \ldots + X_n$. Then $E[X_i] = (1-p)^{n-1}$ so by linearity of

At the other extreme $i = k-1$

$$g(k-1) = \frac{k(n-k)2^{-(k-1)}}{\binom{n}{k}2^{-\binom{k}{2}}} \sim \frac{2kn2^{-k}}{E[X]}$$

As $k \sim 2\log_2 n$ the numerator is $n^{-1+o(1)}$. The denominator approaches infinity and so $g(k-1) = o(1)$. Some detailed calculation (which we omit) gives that the remaining $g(i)$ are also negligible so that Corollary 1.3.5 applies. □

Theorem 1.1 leads to a strong concentration result for $\omega(G)$. For $k \sim 2\log_2 n$

$$\frac{f(k+1)}{f(k)} = \frac{n-k+1}{k+1}2^{-k} = n^{-1+o(1)} = o(1)$$

Let $k_0 = k_0(n)$ be that value with $f(k_0) \geq 1 > f(k_0+1)$. For "most" n the function $f(k)$ will jump from a large $f(k_0)$ to a small $f(k_0+1)$. The probability that G contains a clique of size k_0+1 is at most $f(k_0+1)$ which will be very small. When $f(k_0)$ is large Theorem 1.1 implies that G contains a clique of size k_0 with probability nearly one. Together, with very high probability $\omega(G) = k_0$. For some n one of the values $f(k_0), f(k_0+1)$ may be of moderate size so this argument does not apply. Still one may show a strong concentration result found independently by Bollobás, Erdős [1976] and Matula [1976].

Corollary 1.2 There exists $k = k(n)$ so that

$$\Pr[\omega(G) = k \text{ or } k+1] \to 1$$

2 Chromatic Number

Again let us fix $p = \frac{1}{2}$ and this time we consider the chromatic number $\chi(G)$ with $G \sim G(n,p)$. Our results in this section will be improved in Lecture 4.

Theorem 2.1. Almost surely

$$\frac{n}{2\log_2 n}(1+o(1)) \leq \chi(G) \leq \frac{n}{log_2 n}(1+o(1))$$

For the lower bound we use the general bound

$$\chi(G) \geq n/\omega(\overline{G})$$

which is true since each color class must be a clique in $\overline{G}$ and so can be used at most $\omega(\overline{G})$ times. But $\overline{G}$ has the same distribution as G so almost surely

Lecture 2: More Random Graphs

1 Clique Number

Now we fix edge probability $p = \frac{1}{2}$ and consider the clique number $\omega(G)$. We set

$$f(k) = \binom{n}{k} 2^{-\binom{k}{2}},$$

the expected number of k-cliques. The function $f(k)$ drops under one at $k \sim 2\log_2 n$. (Very roughly, $f(k)$ is like $n^k 2^{-k^2/2}$.)

Theorem 1.1 Let $k = k(n)$ satisfy $k \sim 2\log_2 n$ and $f(k) \to \infty$. Then almost always $\omega(G) \geq k$.

Proof. For each k-set S let A_S be the event "S is a clique" and X_S the corresponding indicator random variable. We set

$$X = \sum_{|S|=k} X_S$$

so that $\omega(G) \geq k$ if and only if $X > 0$. Then $E[X] = f(k) \to \infty$ and we examine Δ^*. Fix S and note that $T \sim S$ if and only if $|T \cap S| = i$ where $2 \leq i \leq k-1$. Hence

$$\Delta^* = \sum_{i=2}^{k-1} \binom{k}{i} \binom{n-k}{k-i} 2^{\binom{i}{2} - \binom{k}{2}}$$

and so

$$\frac{\Delta^*}{E[X]} = \sum_{i=2}^{k-1} g(i)$$

where we set

$$g(i) = \frac{\binom{k}{i}\binom{n-k}{k-i}}{\binom{n}{k}} 2^{\binom{i}{2}}$$

Observe that $g(i)$ may be thought of as the probability that a randomly chosen T will intersect a fixed S in i points times the factor increase in $\Pr[A_T]$ when it does. Setting $i = 2$,

$$g(2) = 2\frac{\binom{k}{2}\binom{n-k}{k-2}}{\binom{n}{k}} \sim \frac{k^4}{n^2} = o(1)$$