Random Graphs and Erdős Magic

Joel Spencer

Asperen 2018

A Special Note: These notes are only an approximation to the actual lectures. There is some extra material here, and there will almost certainly be some material in the lectures not covered in the notes. In particular, the split into specific days should be regarded only as an estimate. – JS

The Probabilistic Method.
Lecture II: Tuesday, October 30

1 Crossing Number

The crossing number κ of a graph G is the least number of crossings of edges when the graph is embedded in the plane.

Theorem 1.1 *Let G have v vertices, e edges and κ crossings. Assume (not the tightest) $e \geq 10v$. Then $\kappa \geq ce^3v^{-2}$. (c an absolute positive constant.)*

We use only that a planar graph on v vertices can have at most $3v-6$ edges. From this, $\kappa \geq e-(3v-6)$. Why? If $e \leq 3v-6$ there is nothing to prove. Otherwise, take a crossing, delete one of the edges creating the crossing, and iterate. To make things less messy lets just use $\kappa \geq e-3v$.

Take a random subset of vertices S of G with each v in S with independent probability p. Call that random graph H. The expected number of vertices of H is vp. The expected number of edges of H is ep^2 as each edge of G remains in H with probability p^2. The expected number of crossings of H is $p^4\kappa$ as each crossing of G remains in H with probability p^4. The expected nuber of crossings minus edges plus 3 times vertices is then $p^4\kappa - p^2e + 3pv$. But we've said that this is always nonnegative. Hence $p^4\kappa - p^2e + 3pv \geq 0$. That is, $\kappa \geq p^{-2}e - 3p^{-3}v$.

Now we use calculus! We set the derivative of the right hand side equal to zero, so $p = \frac{9v}{2e}$. The calculus doesn't appear in the formal proof, we simply set p to this value. As p is a probability we need $p \leq 1$ which leads to the side condition $e \geq 4.5v$. Plugging in this value give $\kappa \geq ce^3v^{-2}$.

`Comment:` This is the correct minimal κ (over all G with v vertices and e edges) up to a constant. Given v, e we can have G as the union of v/t disjoint complete graphs on t vertices, with $t \sim c_1e/v$. Each complete graph would give at most t^4 crossing and so the total number of crossings would be at most $(v/t)t^4 \sim c_2 e^3 v^{-2}$.

Asymptopia of $m(n)$

Recall the theorem proven in previous notes.

Theorem 1.2 *(Erdős 1964) If there exists v with*

$$2^v \left(1 - \frac{2\binom{v/2}{n}}{\binom{v}{n}}\right)^m < 1,$$

then $m(n) \leq m$, i.e., there is a family $A_1, A_2, \ldots, A_m \subset \{1, \ldots, v\}$ with no 2-coloring.

We wish to solve for m and then choose v to maximize m. We begin by using the inequality $1 - p \leq e^{-p}$ to find

$$m \geq \left\lceil \frac{v \ln 2}{2\binom{v/2}{n} / \binom{v}{n}} \right\rceil$$

Define

$$p \equiv \frac{2\binom{v/2}{n}}{\binom{v}{n}}$$

We need to explore the asymptotics of p. To gain some intuition about how tight a bound we need, recall what the quantity p represents. In our original proof, p was the probability that a randomly chosen set from a two-colored v-sized universe is monochromatic. This is equivalent to the probability in the following experiment: Fill a bag with $\frac{v}{2}$ red balls and $\frac{v}{2}$ blue balls. Draw n balls at random without replacement. Then p is the probability that all n balls have the same color. One might be tempted to bound p using the approximation $\binom{n}{k} \approx \frac{n^k}{k!}$. This approximation yields $p = 2^{1-n}$, which is the same as the probability of drawing n monochromatic balls with replacement. Intuitively it is clear that for small v, drawing with replacement and drawing without replacement yield different results. As we are trying to minimize m (and thus will be dealing with small v), we look for a better approximation.

First we rewrite p.

$$\begin{aligned} p &= \frac{2\binom{v/2}{n}}{\binom{v}{n}} \\ &= 2 \prod_{i=0}^{n-1} \frac{v/2 - i}{v - i} \\ &= 2^{1-n} \prod_{i=0}^{n-1} (1 - \frac{i}{v - i}) \\ &= 2^{1-n} \exp(\sum_{i=0}^{n-1} \ln(1 - \frac{i}{v - i})) \end{aligned}$$

By the Taylor expansion,

$$\ln(1-\epsilon) \sim -\epsilon - \frac{\epsilon^2}{2} - \dots$$

Therefore, for (see later for what happens when this doesn't hold) $v >> n^{1.5}$,

$$\ln(1 - \frac{i}{v-i}) \sim \frac{-i}{v}$$

Note we can safely ignore the second order term of the Taylor expansion because $\sum_{i=0}^{n} i^2 \sim n^3$ and, by assumption, $v^2 >> n^3$, so the $(\frac{i}{v-i})^2$ term will tend to zero after the summation.

This gives our final approximation for p,

$$p \sim 2^{1-n} \exp(\frac{-n^2}{2v})$$

from where

$$m(n) \leq \lceil 2^{n-1} (\ln 2) v e^{n^2/2v} \rceil$$

for $v >> n^{1.5}$. Let $z = ve^{n^2/2v}$. We want to find v that minimizes z. Using standard calculus techniques,

$$y \equiv \ln z = \ln v + \frac{n^2}{2v}$$

$$y' = \frac{1}{v} - \frac{n^2}{2v^2} = 0$$

giving

$$v = \frac{n^2}{2}$$

for a final result

$$m(n) \leq \left\lceil \frac{e \ln 2}{4} n^2 2^n \right\rceil$$

While this is the correct answer we haven't completed our rigorous argument. We assumed $v \gg n^{1.5}$ in making our approximations. We now need to check that if v is not that large then the value is bigger than $cn^2 2^n$. This is a somewhat typical (and annoying) part because here the value will be *very* much bigger than $n^2 2^n$ – so we can use crude tools. (Often in asymptotics the calculations are easier when the results are tighter.) Lets take, leaving room, $v \leq n^{1.51}$. The value of p is (remember that n is fixed) an increasing function of v (check the representation as a product) and so it suffices to look at $v = n^{1.51}$. But there the approximation $p \sim 2^{1-n} \exp(-n^2/2v)$ does hold and this gives a p which is *exponentially* small and so a bound on m which is exponentially large, so it is *way* off from the real value.

Asymptopia of $\binom{n}{k}$

In many of the problems in this field, we are faced with an $\binom{n}{k}$ that we need to approximate. There is no single approximation for $\binom{n}{k}$ which always works. Rather, the approximation one should use depends greatly on the relationship between n and k. Here we discuss several approximations of $\binom{n}{k}$ for various relationships between n and k.

First, if both n and k are fixed, $\binom{n}{k}$ has a definite value. Use it. Also, if just k is fixed, clearly $\binom{n}{k} \sim \frac{n^k}{k!}$ is a good approximation.

When both n and k grow, things get more complicated. First, we consider the case where $k = o(n)$. Notice

$$\begin{aligned} \binom{n}{k} &= \frac{(n)_k}{k!} \\ &= \frac{n^k}{k!} \prod_{i=0}^{k-1} (1 - \frac{i}{n}) \\ &= \frac{n^k}{k!} \exp(\sum_{i=0}^{k-1} \ln(1 - \frac{i}{n})) \end{aligned}$$

Using Stirling's formula, we can approximate $k! \sim k^k e^{-k} (2\pi k)^{\frac{1}{2}}$, and n^k is fine as is. What we have left to deal with is the $\ln(1 - \frac{i}{n})$ term. Here we can use the Taylor expansion

$$\ln(1 - \epsilon) = -\epsilon - \frac{\epsilon^2}{2} - O(\epsilon^3)$$

from where

$$\sum_{i=0}^{k-1} \ln(1 - \frac{i}{n}) = \frac{-k^2}{2n} - \frac{k^3}{6n^2} - O(\frac{k^4}{n^3})$$

Of course, we can retain even more terms of the Taylor expansion if we need to. For each extra term we get yet another estimate that generalizes the previous estimate.

The first three terms listed above yeild the following approximations:

- For $k = o(n^{1/2})$, $\sum_{i=0}^{k-1} \ln(1 - \frac{i}{n}) \sim 0$.
- For $k = o(n^{2/3})$, $\sum_{i=0}^{k-1} \ln(1 - \frac{i}{n}) \sim \frac{-k^2}{2n}$.
- For $k = o(n^{3/4})$, $\sum_{i=0}^{k-1} \ln(1 - \frac{i}{n}) \sim \frac{-k^2}{2n} + \frac{-k^3}{6n^2}$.

If $k = o(n)$, we have a useful logarithmically asymptotic result. Since $k! \geq (\frac{k}{e})^k$ for all k,

$$\binom{n}{k} \leq \frac{n^k}{k!} \leq (\frac{ne}{k})^k$$

Furthermore, for some computable constant c,

$$\binom{n}{k} \geq \frac{(n-k)^k}{k!} \geq (\frac{ne}{k})^k c k^{-1/2} (1 - \frac{k}{n})^k$$

where $ck^{-1/2}(1 - \frac{k}{n})^k = n^{o(k)}$. Combining these inequalities, we have the result

- For $k = o(n)$, $\binom{n}{k} = (\frac{ne}{k})^{k(1+o(1))}$.

If $k = \alpha n$ for some $\alpha \in (0,1)$, we can use Stirling's formula to see

$$\begin{aligned}\binom{n}{\alpha n} &= \frac{n!}{(\alpha n)!((1-\alpha)n)!} \\ &\sim \frac{n^n e^{-n}(2\pi n)^{\frac{1}{2}}}{(\alpha n)^{(\alpha n)} e^{-(\alpha n)}(2\pi\alpha n)^{\frac{1}{2}}((1-\alpha)n)^{(1-\alpha)n} e^{-(1-\alpha)n}(2\pi(1-\alpha)n)^{\frac{1}{2}}} \\ &= \frac{n^{-\frac{1}{2}}(2\pi\alpha(1-\alpha))^{-\frac{1}{2}}}{(\alpha)^{\alpha n}(1-\alpha)^{(1-\alpha)n}}\end{aligned}$$

For convenience, let $c_\alpha \equiv (2\pi\alpha(1-\alpha))^{-\frac{1}{2}}$. We use the entropy function $H(\alpha)$ defined as

$$H(\alpha) \equiv -\alpha \log \alpha - (1-\alpha)\log(1-\alpha)$$

One sets $H(0) = H(1) = 0$ so that this is continuous on $[0,1]$. H is symmetric about $\alpha = \frac{1}{2}$ and assumes its maximum at $\alpha = \frac{1}{2}$ with $H(1/2) = 1$. The slope of $H(\alpha)$ at 0 and 1 is ∞ and $-\infty$. We can estimate $H(\frac{1}{2}+\epsilon)$ as $H(\frac{1}{2}) + H'(\frac{1}{2})\epsilon + H''(\frac{1}{2})\frac{\epsilon^2}{2} + O(\epsilon^3) = 1 - \frac{4}{\ln 2}\epsilon^2 + O(\epsilon^3)$.

Now we can succinctly state our result.

- For $k = \alpha n$, $\binom{n}{\alpha n} \sim c_\alpha n^{-\frac{1}{2}} 2^{nH(\alpha)}$.

Improving bounds of $m(n)$

In 1963 Erdős gave a lower bound for $m(n)$ of 2^{n-1}. In 1964, he gave an upper bound of $\frac{e\ln 2}{4}n^2 2^n$. These bounds are quite far apart, especially if you consider 2^n as 1. In 1978 Beck improved Erdős's result giving a lower bound of $2^n n^{1/3}$. In 2000 Radhukrishnan and Srinivasan improved the lower bound to $2^n(\frac{n}{\ln n})^{1/2}$. There is, of course, still room for improvement, but these are the best known bounds to date. We will give the Radhukrishnan Srinivasan bounds using a beautiful new argument of Kozik and Cherkashin.

It will be convenient to parametrize $m = 2^{n-1}k$. Let an arbitrary family of $m = 2^{n-1}k$ sets, each of size n. Let Ω be the set of elements. (Note: the size of Ω is arbitrary.)

Kozik and Cherkashin give a *Random Greedy* algorithm to color Ω. Using Erdős Magic they only need show that the probability the algorithm fails is less than one. Turning things around, at the end they make k as big as possible so that the failure probability (well, the upper bound on the failure probability) is less than one.

The random part consists of randomly ordering Ω. To facilitate the analysis to each $v \in \Omega$ we assign an independent uniform "birthtime" $t_v \in [0,1]$. (With probability one no two birthtimes are the same.) Begin at time zero with no vertices colored. At time t_v we decide how to color v by the following rediculously easy protocol:

`The Protocol:` Color v Red **unless** doing so would create a totally red set. When that is the case color v Blue.

Let BAD be the event that this randomized algorithm produces a monochromatic set.

Tautologically, a Red set cannot be created. Suppose a blue set A_i is created. Let v be that point of A_i with the least birthtime. Why wasn't v colored red? There must have been a set A_j (there may be more than one) such that making v red would have made A_j red. When this happens we say A_i *blames* A_j and we denote this event by $BLAME[A_i, A_j]$. There are less than $(2^{n-1}k)^2$ choices of A_i, A_j so

What is the probability of $BLAME[A_i, A_j]$. First of all, A_i, A_j must intersect in precisely one vertex, call it v. Set $t = t_v$, the birthtime of v, which is uniform over $[0, 1]$. All other $w \in A_i$ must have birthdates in $[t, 1]$ and all other $z \in A_j$ must have birthdays in $[0, t]$. As birthdates are independent this has probability $(1-t)^{n-1}t^{n-1}$. Thus $BLAME[A_i, A_j]$ has probability at most $\int[t(1-t)]^{n-1}$. There are less than $(2^{n-1}k)^2$ pairs A_i, A_j so

$$\Pr[BAD] \leq (2^{n-1}k)^2 \int [t(1-t)]^{n-1}$$

This leads to a nice asymptotic calculus problem, what is the maximal k (asymptotic in n) so that the right hand side is ≤ 1?

But it turns out (it took several years to see this!) that one can do a better analysis without changing the algorithm. Let $\epsilon = \epsilon(n, k)$. Call a set $A = A_i$ *unbalanced* if either the largest birthtime t_v amongst the $v \in A$ is less than $(1-\epsilon)/2$ or the least birthtime t_v amongst the $v \in A$ is greater than $(1+\epsilon)^n$. Then A is unbalanced with probability $2 \cdot ((1-\epsilon)/2)^n$. Let $BAD2$ be that some A is unbalanced. Then

$$\Pr[BAD2] \leq 2^{n-1}k \cdot 2 \cdot ((1-\epsilon)/2)^n = k(1-\epsilon)^{n-1}$$

Suppose no A is unbalanced. Then for A_i to blame A_j it must be that their common vertex v has birthdate $t = t_v$ within $\epsilon/2$ of $1/2$. This occurs with probability ϵ. Given the birthdate the probability is $[t(1-t)]^{n-1}$ which we bound by $4^{-(n-1)}$. Let $BAD1$ be that no A is unbalance and yet $BLAME[A_i, A_j]$ for some i, j. Then

$$\Pr[BAD1] \leq (2^{n-1}k)^2 \epsilon 4^{-(n-1)} = k^2\epsilon$$

For BAD to occur we must have $BAD1$ or $BAD2$. Thus

$$\Pr[BAD] \leq k(1-\epsilon)^{n-1} + k^2\epsilon$$

Another calculus problem! What is the largest $k = k(n)$ such that there exists $\epsilon = \epsilon(k, n)$ with $k(1-\epsilon)^{n-1} + k^2\epsilon \leq 1$. Asymptotic answer: Select $\epsilon = C(\ln n)/n$ to kill off the first term and then $k = c_1\epsilon^{-1/2} = c_2\sqrt{n/(\ln n)}$ will make the second addend small.

Many of the more recent (and more exciting) results involve the analysis of Random Processes or Randomized Algorithms. Let us even give this a name:

> **Modern Erdős Magic:** If there is a randomized algorithm that creates an object with desired properties with positive proability that that object must exist.