The story of $R(3, k)$ and random graph processes
Lecture Notes
The Probabilistic Method, SS 2016

Tibor Szabó

These are essentially the notes I prepared as a background material in Combinatorics and Discrete Probability for the (Pre-)Doc-Course we held a few years ago about the so-called triangle-free process (the analysis of which goes way beyond the ambitions of our present course). Comments are welcome.

1 The first moment method in combinatorics

1.1 Ramsey numbers

It all starts with the Pigeon Principle: if each of n pigeons sits inside one of k pigeonholes, then there should be a hole with at least $\lceil n/k \rceil$ pigeons sitting in it. Despite its simplicity the Pigeonhole Principle is a powerful tool for proving existence, in fact probably the earliest appearance of the probabilistic method.

The simplest form of the Pigeonhole Principle was formulated by Dan Kleitman: "Of three ordinary people two must have the same sex." To say this slightly more generally: in every `red`/`blue`-coloring of the *vertices* of the complete graph K_{k+l-1} either there is a `red` K_k or a `blue` K_l. And this statement is of course tight, as coloring $k-1$ vertices of K_{k+l-2} `red` and $l-1$ `blue` shows.

Coloring the edges of the complete graph instead of its vertices already makes the problem significantly harder. A standard first semester combinatorics exercise asks us to show that in every party of six there are either three people who mutually know each other or three people who mutually do not know each other. Then one models the problem with coloring the edges of K_6: vertices are the people, `red` edge corresponds to knowing each other, `blue` corresponds to not knowing each other and one shows that there is a monochromatic clique (a clique whose edges all have the same color) of order 3.

In general we define the Ramsey number

$$R(k, l) = \min\{N : \forall c : E(K_N) \to \{\texttt{blue},\texttt{red}\} \exists \texttt{ red } K_k \text{ or } \texttt{blue } K_l \}$$

The exercise above asked us to prove $R(3, 3) \leq 6$.

The classic theorem Ramsey proves the finiteness of the Ramsey number.

Theorem 1 *(Ramsey, 1930) For every $k, l \geq 1$, $R(k, l) < \infty$*

The exact values of the Ramsey numbers are not known, even for very small values a computer search fails. The largest symmetric value known is $R(4,4) = 18$ and this is a homework exercise. The value of $R(5,5)$ is somewhere between 43 and 49. The value of $R(3,l)$ is known up to $l \leq 9$, most recently it was determined that $R(4,5) = 25$ and that's it. These are all precise values known. But we are not really interested in the small values, rather want to understand the asymptotic behaviour.

Theorem 2 *(Erdős-Szekeres, 1935)* $R(k,l) \leq R(k,l-1) + R(k-1,l)$

Proof: Take $N = R(k,l-1) + R(k-1,l)$ and an arbitrary `red`/`blue` coloring of $E(K_N)$. Pick an arbitrary vertex $x \in V$. One of the following two cases happens: *Case 1*: x has at least $R(k-1,l)$ `red` neighbors or *Case 2*: x has at least $R(k,l-1)$ `blue` neighbors. In Case 1: if there is a `red` K_{k-1} among the red neighbors of x, then together with x they form a `red` K_k, done. Otherwise there is a `blue` K_l among the red neighbors of x and we are also done.
Case 2 is analogous. □

Corollary 1 *For all* $k, l \geq 1$, $R(k,l) \leq \binom{k+l-2}{k-1}$. *In particular,* $R(k,l)$ *exists.*

Proof. Induction on $k + l$. □

This immediately implies the following exponential upper bound.

Corollary 2 $R(k,k) \leq 4^k$.

How about a lower bound?

After a first encounter with the problem, the coloring most of us would come up with after five minutes (and most likely even after five months ...) is the Turán-coloring: Partition $(k-1)^2$ vertices into parts of size $k-1$ and color each edge within parts by `red` and edges between parts with `blue`. The largest monochromatic clique has size $k-1$, proving $R(k,k) \geq (k-1)^2 + 1$. Pretty weak considering that the upper bound is exponential. Unless of course the Turán coloring was optimal, which was the conjecture for some years in the 1940's.

An attractive alternative coloring discussed in the exercises is the Paley-coloring.

Definition (of *Paley-coloring*) Let $p \equiv 1 \pmod 4$ (this congruence assumption is needed for the coloring to be well-defined). Label the vertices of K_p with $\mathbb{F}_p$ (field of p elements) and color an edge xy `red` iff $x - y \in Q_p$, where $Q_p = \{z^2 : z \in \mathbb{F}_p\}$ is the set of *quadratic residues* modulo p.

The `red` graph for $p = 5$ is the 5-cycle. It does not contain a monochromatic K_3, providing the lower bound for $R(3,3) = 6$. In the exercises we will show that the Paley-coloring does not contain a monochromatic K_4 for $p = 17$. How good the Paley-colorings are in general? It is not known. Numerical data suggests that the largest monochromatic clique might be much smaller than the square root of the number of vertices. For example, for $p = 6997$ the clique number is only 17 (by computer calculations of Shearer). In general, however, it is only known that the largest monochromatic clique is of the order $\sqrt{p}$, giving the same meager $\Omega(k^2) = R(k,k)$ as the Turán-coloring. Any improvement on the Paley-coloring would represent a major breakthrough in number theory.

Is there a super-quadratic construction for the symmetric Ramsey problem? In the Paley-coloring every vertex has the same number of `red` and `blue` neighbors, since $|Q_p| = (p-1)/2$. After some more work with the eigenvalues of its adjacency matrix one can also show that any constant r vertices have roughly $p/2^r$ vertices in their common `red` neighborhoods and further random-like properties hold as well.

The following fundamental result of Erdős is considered the birth of the probabilistic method.

Theorem 3 *(Erdős, 1947)* $R(k,k) \geq \sqrt{2}^k$

Proof. Take a random `red`/`blue`-coloring c of the edges of K_N, where each edge is colored `red` with probability $1/2$, and these events are all mutually independent from each other. For a fixed k-subset S of the vertices of let A_S be the event that the edges inside S all have the same color under c. Then $Pr(A_S) = \frac{2}{2^{\binom{k}{2}}}$. We do some seemingly crude estimations to bound the probability of the event that there exists a monochromatic k-subset:

$$Pr\left(\bigvee_{S\in\binom{V}{k}} A_S\right) \leq \sum_{S\in\binom{V}{k}} Pr(A_S) = \binom{N}{k}2^{1-k} \leq 2\cdot\left(\frac{eN}{k2^{(k-1)/2}}\right)^k \to 0,$$

provided $N = \left(\frac{1}{e\sqrt{2}} - \epsilon\right)k\sqrt{2}^k$ where $\epsilon > 0$ is an arbitrary constant.

Hence for such an N there must exist a two-coloring without a monochromatic k-clique, proving the lower bound. □

Note that the following proof also gives that *most* of the two-colorings are good for our Ramsey-purpose. However nobody is able to *construct* explicitly colorings on 1.000001^k vertices without a monochromatic k-clique. An unexpected idea was needed just to construct explicitly a coloring on a super-quadratically many vertices. (Nagy, Frankl-Wilson, ...)

A \$1000 dollar question of Erdős, still wide open: Determine $\lim_{k\to\infty}\sqrt[k]{R(k,k)}$. Proving the existence of the limit is worth \$500. The upper bound was improved recently by David Conlon to $\frac{4^k}{p(k)}$, where $p(k)$ is a function tending to infinity faster than any polynomial. Still, nobody knows to prove $R(k,k) \leq 3.999^k$. The lower bounds stands firmly where it essentially was in 1947: $R(k,k) = \Omega(k\sqrt{2}^k)$.

In the proof of Theorem 3 we used a couple of standard estimates; we take the opportunity to list the ones we use most frequently.

The triviality of estimating the probability of the union of events with the sum of their probabilities is called the **union bound**.

1.1.1 Estimates and asymptotics

- Stirling formula: $n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$
- Binomial coefficients: $\left(\frac{n}{k}\right)^k \leq \binom{n}{k} \leq \left(\frac{ne}{k}\right)^k$
- Almost 1: $e^{-\frac{x}{1-x}} \leq 1 - x \leq e^{-x}$

Our **Asymptotic notation:** Let $f, g : N \to R$ be two functions.

- $f = O(g)$ if there exist constants C and K such that $|f(n)| \leq C|g(n)|$ for $n \geq K$
- $f = \Omega(g)$ if $g = O(f)$
- $f = \Theta(g)$ (or sometimes we write $f \sim g$) if $f = O(g)$ and $f = \Omega(g)$
- $f = o(g)$ (or sometimes we write $f \ll g$) if $\lim_{n\to\infty} \frac{f(n)}{g(n)} = 0$
- $f \approx g$ if $\lim_{n\to\infty} \frac{f(n)}{g(n)} = 1$

1.2 Turán's Theorem

In a typical application of the union bound we create a probability space over a set of objects and we prove that with non-zero probability a random object from this space has the property $\mathcal{P}$ we are interested in. And then, we make the earth-shaking conclusion that **there exists** an object possessing the required property. Not only sometimes, or with probability 1% or 99%, but absolutely surely, with **100% probability**, THERE EXISTS such an object.

Already the union bound gives many surprising results, which are not known to have a proof in any other way. A slightly more general form of it is to use the **linearity of expectation**. If $X_1, \ldots, X_k$ are random variables then

$$E(X_1 + \cdots + X_k) = E(X_1) + \cdots + E(X_k).$$

The trivial, yet extremely useful conclusion we can immediately draw from knowing that the expected number of a random variable X is 11 for example is that, (again **for sure**, **without any doubt**), **there is** an object in our space whose X-value is *at least* 11. And we also know that there is an object whose X-value is *at most* 11.

To present it at work we prove the classic Theorem of Turán.

Theorem 4 *For an arbitrary graph G we have*

$$\alpha(G) \geq \sum_{v\in V(G)} \frac{1}{d(v)+1}$$

Proof. Choose an ordering $<$ of the vertices uniformly at random among all $n!$ possible orderings.

Define the random set

$$I = I(<) = \{v \in V : v < u \text{ for all } u \in N(v)\}.$$

Then I is independent since if $xy \in E$ and, say, $x < y$, then y is not in I, because it has a neighbor (namely x), which comes *before it* in the ordering $<$.

Now we just have to prove that I is big. Let X_v be the characteristic random variable of the event that vertex v is in the set I. Then

$$E(X_v) = Pr(v \text{ precedes all its neighbors according to } <) = \frac{1}{d(v)+1}.$$

For the size of I we have that $|I| = \sum_{v \in V} X_v$, so by the linearity of expectation

$$E(|I|) = \sum_{v \in V} \frac{1}{d(v)+1}.$$

Hence there exists an ordering $<$ such that the independent set I defined with its help has the appropriate size. □

Corollary 3 *(Turán's Theorem, 1941) The number of edges in any K_{k+1}-free graph is at most $\left(1 - \frac{1}{k}\right)\frac{n^2}{2}$.*

Proof. Let G be a K_{k+1}-free graph. We apply the previous theorem for the complement of G. Since $x \to \frac{1}{x+1}$ is a convex function, we have that

$$k \geq \alpha(\bar{G}) \geq \sum_{v \in V(G)} \frac{1}{n - d_G(v)} \geq \frac{n}{n - \frac{2e(G)}{n}},$$

and the bound follows. □

For simplicity we derived only the asymptotic version of Turán's Theorem; with a bit more care the full result, characterizing the extremal graph can also be obtained.

2 Initial bounds for $R(3, k)$

Now we return to the Ramsey problem and discuss a bit the asymmetric version. $R(2, k)$ is of course k for every k. The first non-trivial case is the value of $R(3, k)$, which will be one of the central problem of the entire Doc-course.

The Erdős-Szekeres bound immediately gives a quadratic upper bound:

$$R(3, k) \leq \binom{k+1}{2} = \Theta(k^2).$$

As usual, we have trouble with the construction. It is unlikely that sitting down and scribbling graphs on a piece of paper one would come up with a construction which gives a super-linear lower bound on $R(3,k)$. From now on we prefer to think of graphs instead of a colorings. By `red` we mean an edge, by `blue` we mean non-edge. We want a triangle-free graph G on n vertices with independence number $\alpha(G)$ as small as possible. Intuitively plausible attempts like thinking that "many edges can only help in reducing the independence number" fail miserably. Indeed, the triangle-free graph with the most number of edges is $K_{\frac{n}{2},\frac{n}{2}}$ and has an independent set of $n/2$; that's a lower bound of $2k$ on $R(3,k)$. The main reason for this failure is that neighborhood of any vertex in a triangle-free graph is an independent set.

Observation. If G is triangle-free, then $\alpha(G) \geq \Delta(G)$.

In order to have a super-linear lower-bound, the degrees must be $o(n)$.

We try with a random construction. The **random graph** $G(n,p)$ is a probability space of graphs on n vertices, where each edge appears with probability $p = p(n)$ and these events are mutually independent. Let us see first, for what edge probability p do we expect to see the first triangle?

For a triple $T = \{a,b,c\} \subseteq V$ let X_T be the characteristic random variable of the event that a and b and c form a triangle. Then $X = \sum_{T \in \binom{V}{3}} X_T$ is the number of triangles in $G(n,p)$. The expectation is $E(X) = \binom{n}{3}p^3 \sim n^3p^3$.

The simple-minded application of the probabilistic methods says that if $E(X) < 1$ we can claim the *existence* of triangle free-graph. Now $E(X) < 1$ if $p \leq \frac{\sqrt[3]{6}}{n}$. How small independence number can we hope for? Well, the expected number of edges is $\binom{n}{2}p \sim \frac{\sqrt[3]{6}}{2}n$. Not so much, the average degree is $\sqrt[3]{6}$. By Turán's Theorem the independence number is at least n over the average degree. In our case this is a linear lower bound; the simple-minded random construction failed.

Here comes the first twist in the probabilistic method idea. When the first triangle appears we could destroy it by deleting a vertex from it. The number of vertices went down by 1, the largest independent set did not grow, their relation is practically unchanged. In fact until the number of triangles is not significant, i.e., not more than say half of the vertices, we can delete one vertex from each, create a triangle-free graph and preserve, up to constant factor the relationship between the independence number and the number of vertices in the graph.

To bound the independence number of the random graph $G(n,p)$ let us calculate the probability that a fixed set S of size r is independent. No pairs must be edges, these events being independent we get $(1-p)^{\binom{r}{2}}$. Hence bounding the union of these events (over all subsets of size r) by the union bound we have

$$Pr(\alpha(G(n,p)) \geq r) \leq \binom{n}{r}(1-p)^{\binom{r}{2}} \leq \left(\frac{ne}{r}\right)^r e^{-p\binom{r}{2}} = \exp(r(\ln n + 1 - \ln r - p\frac{r-1}{2}))$$

The expression in the exponent is at most $-r(\ln r - 1)$ provided $r \geq \frac{2\ln n}{p} + 1$. So the probability that $\alpha(G) \geq \frac{2\ln n}{p} + 1$ tends to 0.

To bound the number T of triangles in $G(n,p)$ from above we can use the expectation and **Markov' Inequality**.

Inequality 1 *(Markov) If $X \geq 0$ almost surely, then for any $t > 0$ we have*

$$Pr(X \geq t) \leq \frac{E(X)}{t}$$

We say that an event E happens *almost surely* and write *a.s.* if $Pr(E) = 1$. For a sequence of events E_n we say they hold *asymptotically almost surely* and write *a.a.s.* if $Pr(E_n) \to 1$.

In our case the expectation is $E(X) = \binom{n}{3}p^3$ hence

$$Pr(X > \frac{n}{2}) \leq \frac{2E(X)}{n} = \frac{(n-1)(n-2)p^3}{3}.$$

If for example $p = n^{-2/3}$, then this probability is less than $\frac{1}{3}$.

Hence there exists an instance of a graph H such that

- there are at most $\frac{n}{2}$ triangles in H and
- $\alpha(H) \leq \lceil \frac{2 \ln n}{p} \rceil = \lceil n^{2/3} \ln n \rceil$

After deleting one vertex from every triangle of H we obtain a triangle-free graph H' with $n' > n/2$ vertices and independence number at most $\Theta(n^{2/3} \ln n)$. This gives a super-linear bound on the asymmetric Ramsey number.

Theorem 5

$$R(3,k) = \Omega\left(\frac{k}{\ln k}\right)^{3/2}.$$

3 An interlude — the second moment method

3.1 Is the lower bound method optimized?

Is there still hope to improve on the analysis of this construction? What if we choose a larger p? Say, just to take an example, $p = n^{-0.6}$. Then the bound on the independence number would be stronger, the largest independent set would have at most $n^{0.6} \ln n$ vertices a.a.s. OK, the expected number of triangles is $\binom{n}{3}p^3 = \Omega(n^{1.2})$ which is far more than n, but maybe, somehow we could still prove that with reasonable probability the number of triangles is not always close to its expectation. We can certainly do a little wishful thinking about a distribution of the number of triangles which would give the required expectation $n^{1.2}$, but still would deviate from it significantly with constant probability.

Unfortunately(?) this is not the case. Such a large deviation from the mean is not likely to happen for X (and nor for most random variables that appear in this course). Markov's Inequality, in its straightforward form, is not able to help us proving that: we need to bound the probability that X is small, that is $Pr(E(X) - t > X)$, where say $t = n^{1.2} - n/2$. We need to use the *second moment* of X in form of **Chebyshev's Inequality**.

Inequality 2 *(Chebyshev) For any $t > 0$ we have*

$$Pr(|X - E(X)| \geq t) \leq \frac{Var(X)}{t^2}$$

Proof. To prove Chebyshev we apply Markov for the nonnegative random variable $(X - E(X))^2$.

$$Pr(|X - E(X)| \geq t) = Pr((X - E(X))^2 \geq t^2) \leq \frac{(E(X - E(X))^2}{t^2} = \frac{Var(X)}{t^2}$$

□

The *variance* $Var(X)$ of a random variable X is defined to be $E((X - E(X))^2) = E(X^2) - E(X)^2$. The square root of the variance is the *standard deviation* $\sigma = \sigma(T)$. Chebyshev's Inequality shows that the probability that a random variable deviates from its expectation by significantly more than the standard deviation tends to 0. If, for example, $f(n)$ is a function tending to infinity arbitrarily slowly then the probability that X takes a value which is at least $f(n)\sigma(X)$ away from the expectation $E(X)$ is at most $\frac{1}{f(n)} \to 0$.

Hence if it so happens that for some random variable X the standard deviation is smaller order than the expectation, then by Chebyshev's Inequality the value of X is concentrated around the expectation a.a.s.

Corollary 4 *If $Var(X) = o(E(X)^2)$ for some random variable X, then*

(i) $Pr(X = 0) \to 0$ *and*

(ii) $Pr(|X - E(X)| \geq \epsilon E(X)) \to 0$ *for every $\epsilon > 0$.*

As we have seen already when $p \ll \frac{1}{n}$, then for the number X of triangles in $G(n, p)$ we have

$$Pr(X \neq 0) \leq E(X) \to 0.$$

When $p \gg \frac{1}{n}$, then $E(X) = \binom{n}{3}p^3 \to \infty$, that's promising, but is it true that $Pr(X \neq 0) \to 1$? Not necessarily if X was just some arbitrary random variable. We calculate the variance. Whenever the variable X is a sum we can write the variance as the sum of covariances.

$$Var(X) = Var(\sum_T X_T) = E((\sum_T (X_T - E(X_T)))^2) = \sum_{T,S} \underbrace{E(X_T X_S) - E(X_T)E(X_S)}_{Cov(X_T, X_S)}$$

Whenever $T = S$, the covariance is equal to $E(X_T) - E(X_T)^2$, since X_T is a characteristic random variable.

If $|T \cap S| \leq 1$ then the the events whether T and S host a triangle are independent, so the covariance is 0.

When $|T \cap S|$ then the events "help each other": T and S both form a triangle with probability p^5 while $E(X_T)E(X_S)$ is just p^6.

It turns out that the contribution of these terms is negligible compared to $E(X)^2$, so the corollary to Chebyshev's Inequality applies. Indeed, the $\binom{n}{3}$ diagonal terms contribute at most p^3 while there are $\binom{n}{3}(n-3)$ other terms contributing at most p^5. Altogether we have that

$$Var(X) \leq n^3p^3 + n^4p^5 = o((n^3p^3)^2).$$

Hence $G(n,p)$ contains a triangle a.a.s. provided $p \gg \frac{1}{n}$. We have proved the following

Theorem 6 *(i)* $\frac{1}{n}$ *is a threshold function for the property that* $G(n,p)$ *contains a triangle.*

(ii) Let $p \gg \frac{1}{n}$. *Then for every* $\epsilon > 0$

$$Pr\left(\left|\# \text{ of triangles in } G(n,p) - \binom{n}{3}p^3\right| \geq \epsilon\binom{n}{3}p\right) \to 0$$

A function $f(n)$ is called a *threshold function* for the property $\mathcal{P}$ if two things hold. On the one hand for $p \ll f(n)$ its is very unlikely that $G(n,p)$ has the property $\mathcal{P}$, on the other hand for $p \gg f(n)$ it is very likely that $G(n,p)$ possesses $\mathcal{P}$. Formally

$$Pr(G(n,p) \text{ has property P}) \to \begin{cases} 0 & \text{if } p \ll f(n) \\ 1 & \text{if } p \gg f(n) \end{cases}$$

3.2 Counting subgraphs

Can we do this subgraph counting for any subgraph G? Let's try first $G = K_4$.
Let now X denote the number of K_4 in $G(n,p)$
The expectation is $E(X) = \binom{n}{4}p^6 \sim n^4p^6$.
The critical probability where the expectation climbs above 1 is at $p = n^{-2/3}$.
In the variance calculation we consider the covariances of X_T and X_S, where T and S are 4-element subsets.
If $|T \cap S| \leq 1$ then the variables are again independent, so the covariance is 0.
When $|T \cap S| = 2$, then the covariance is $p^{11} - p^{12} \leq p^{11}$,
when $|T \cap S| = 3$, the covariance is $p^9 - p^{12} \leq p^9$,
while if $T = S$, the covariance is $E(X_T) - E(X_T)^2 \leq p^6$.
There are $\binom{n}{4}\binom{n-4}{2} \sim n^6$ terms of the kind $|T \cap S| = 2$, and $\binom{n}{4}(n-4) \sim n^5$ terms

of the kind $|T \cap S| = 3$, and $\binom{n}{4} \sim n^4$ terms of the kind $T = S$. In conclusion the variance is at most

$$Var(X) \leq n^4p^6 + n^4p^{11} + n^5p^9 = o((n^4p^6)^2)).$$

The last equality is valid since each of the three terms on the left hand side is dominated by n^8p^{12}, provided $p \gg n^{-2/3}$ (or equivalently if $E(X) \sim n^4p^6 \to \infty$). In conclusion, the corollary to Chebyshev's Inequality applies and a theorem analogous to Theorem 6 holds for K_4. The threshold probability of the appearance of K_4 in $G(n,p)$ is $n^{-2/3}$ and above this probability the variable X is concentrated about its expectation.

The threshold functions for the appearance of K_3 and K_4 and their analysis readily suggests that n^{-v_G/e_G} is the threshold function for the appearance of any G in $G(n,p)$. This is true (and was confirmed by Erdős and Rényi around 1960 in their classic papers initiating random graph theory) for a large class of graphs, but not arbitrarily. For example, if G is the 4-clique with one extra edge attached to one of its vertices, then it is *not* true that for $p \gg n^{-v_G/e_G} = n^{-5/7}$ the random graph $G(n,p)$ contains G a.a.s. The reason for this that, trivially,

$$Pr(G(n,p) \supseteq G) \leq Pr(G(n,p) \supseteq K_4) \leq \binom{n}{4}p^6 \to 0,$$

if, for example $p = n^{-20/29} \ll n^{-2/3}$. Since $n^{-20/29} \gg n^{-5/7}$ we have a problem. In fact, the problem is quite visible for the particular G: it has a much "denser" subgraph (namely K_4) than itself. This subgraph appears in $G(n,p)$ only at a much larger probability than the one we expected G to appear at, but its presence is of course necessary for G to appear. Hence the correct threshold for G should be the one where its densest subgraph appears.

This motivates the following definition.

$$m(G) = \max\left\{\frac{e_H}{v_H} : H \subseteq G, v_H > 0\right\}$$

A graph G is called *balanced* if $m(G) = \frac{e_G}{v_G}$ and it is called *strictly balanced* if $m(G) > \frac{e_H}{v_H}$ for all $H \subset G$. For example every clique is strictly balanced. (HW: How about C_n? and the cube Q_n?)

Theorem 7 *(Bollobás, 1981) Let G be a graph with at least one edge. Then*

(i) $n^{-\frac{1}{m(G)}}$ *is a threshold function for the property that $G(n,p)$ contains a copy of G.*

(ii) Let $p \gg n^{-\frac{1}{m(G)}}$ be bounded away from 1. Then $X_G \sim E(X_G)$.

Remark. This theorem was proved by Erdős and Rényi for balanced G much earlier.

Proof. The proof of this statement is a more or less a repetition of what was said before for K_3 and K_4, but the choice of a good notation could make a difference. Let $H^* \subseteq G$ be a subgraph with $\frac{e_{H^*}}{v_{H^*}} = m(G)$. Let us assume first that $p \ll n^{-1/m(G)} = n^{-v_{H^*}/e_{H^*}}$. Then

$$\begin{aligned} Pr(G(n,p) \supseteq G) &\leq Pr(G(n,p) \supseteq H^*) \\ &\leq E(X_{H^*}) = \frac{n(n-1)\cdot \dots \cdot (n - v_{H^*} + 1)}{|Aut(H^*)|} p^{e_{H^*}} \\ &\leq n^{v_{H^*}} p^{e_{H^*}} = o(1). \end{aligned}$$

In calculating the expectation we count the number of copies of H^* in K_n: there are $n(n-1)\cdot \dots \cdot (n - v_{H^*} + 1)$ ways to place the vertices in a labelled fashion, however each automorphism will permute these vertices in such a way that the edges are not affected, leading to the same "copy" of G in K_n.

Suppose now that $p \gg n^{-1/m(G)} = n^{-v_{H^*}/e_{H^*}}$. Then clearly $E(X_G) \to \infty$. Let $X_G = \sum I_C$, where the summation runs through all copies C of G in K_n and I_C is the characteristic random variable that the copy C is present in $G(n,p)$. Clearly, $E(I_C) = p^{e_G}$. We calculate the variance

$$\begin{aligned} Var(X_G) = \sum_{C_1, C_2} Cov(I_{C_1}, I_{C_2}) &= \sum_{|E(C_1)\cap E(C_2)|\geq 1} E(I_{C_1} I_{C_2}) - E(I_{C_1})E(I_{C_2}) \\ &= \sum_{H \subseteq G} \sum_{C_1 \cap C_2 \simeq H} p^{2e_G - e_H} - p^{2e_G} \\ &\leq \sum_{H \subseteq G} n^{2v_G - v_H} (p^{2e_G - e_H} - p^{2e_G}) \\ &\leq \sum_{H \subseteq G} \frac{n^{2v_G} p^{2e_G}}{n^{v_H} p^{e_H}} \\ &\sim \sum_{H \subseteq G} \frac{E(X_G)^2}{n^{v_H} p^{e_H}} = o(E(X_G)^2) \end{aligned}$$

The last equality holds since there are constantly many summands (as many as there are subgraphs in G) and the denominator in each of them tends to ∞. Indeed, $n^{v_H} p^{e_H} = (np^{e_H/v_H})^{v_H} \to \infty$ since $np^{e_H/v_H} \geq np^{e_{H^*}/v_{H^*}} \to \infty$.

Hence the corollary to Chebyshev's Inequality applies and we are done. □

4 Improving the lower bound

Let us return now to our quest to improve the lower bound on $R(3,k)$. A reminder: We proved the existence of a triangle-free graph with independence number $\sim n^{2/3}\log n$. We used what is called in probabilistic combinatorics the **alteration** method or **deletion method**: first we constructed a random object with certain property (small independence number) and then deterministically modified it (deleted those vertices which were contained in a triangle).

We also convinced ourselves with the second moment method that there is no hope of improving this construction idea: the number of triangles will concentrate about its mean and since the mean does go above n (the tolerance level of our approach) immediately when we increase the probability to exceed $n^{-2/3}$, and hence we might be removing all vertices.

Removing vertices to make a graph triangle-free is quite a barbaric surgery: there is only this local cancerous part, a triangle, cutting out a vertex of it damages much more than necessary, many edges that were not part of a triangle are also removed. Indeed, at the critical probability $p = n^{-2/3}$ the number of triangles is around $\binom{n}{3}p^3 \sim n$ while the number of edges is expected to be $\binom{n}{2}p \sim n^{4/3}$. So only a negligible fraction of the edges is contained in a triangle.

For an improved construction we want to remove from $G(n,p)$ only what's necessary: an *edge* from each triangle. More precisely, we consider $G(n,p)$ and take an arbitrary maximal family $\mathcal{T}$ of edge-disjoint triangles and remove all edges of the members of $\mathcal{T}$. By definition every triangle intersects some member of $\mathcal{T}$ in at least one edge, so after the deletion of these edges, the graph is triangle-free. How large is now the largest independent set? How do we analyse this?

As a start let's determine the largest probability p at which we have a chance to succeed. The expected number of K_3 takes over the expected number of edges when $\binom{n}{3} \sim \binom{n}{2}$, that is when $p \sim n^{-1/2}$. Soon above this threshold we have no chance to succeed: the number of triangles becomes much larger than the number of edges and soon (another log-factor above) all edges will participate in a triangle, so we would need to delete all of them to have a triangle-free graph. Hence we set our

$$p = \epsilon n^{-1/2},$$

then the number of triangles is $\approx \frac{\epsilon^3}{6}n^{3/2}$ and the number of edges is $\approx \frac{\epsilon}{2}n^{3/2}$, much more. Here $\epsilon > 0$ is a fixed small constant to be determined later.

We already proved that in every subset of size $\frac{2\log n}{p}$ there is at least one edge a.a.s. We now want to show that for just slightly larger k, say when $k > \frac{17\log n}{p}$ not only there is at least one edge in every k-set, but also there are many, at least half of the expectation $\frac{1}{2}\binom{k}{2}p$. Let us fix a set $S, |S| = k$ and let Z_S be the random variable that counts the number of edges in S. Then if we use Chebyshev's Inequality

$$Pr(S \text{ contains at most } \tfrac{1}{2}\tbinom{k}{2}p \text{ edges}) \leq Pr(|Z_S - E(Z_S)| \geq \frac{1}{2}E(Z_S)) \leq \frac{4Var(Z_S)}{E(Z_S)^2}.$$

What is the variance? The random variable Z_S is the sum of indicator random variables I_e for each pair $e \subseteq S$. For this, the variance calculation is much easier, since these indicator variables are mutually independent, so all covariances are 0, except the diagonal ones.

$$Var(Z_S) = \sum_{e \subseteq S} E(I_e) - E(I_e)^2 = \binom{k}{2}(p - p^2) \approx E(Z_S),$$

so

$$Pr(S \text{ contains at most } \tfrac{1}{2}\tbinom{k}{2}p \text{ edges}) \lessapprox \frac{4}{E(Z_S)} \to 0.$$

Great. For every set S, the probability that they have many edges tends to 0. The only problem is that we want to guarantee this simultaneously for ALL subsets of size k. And the number of these subsets is huge, a super-polynomial $\binom{n}{k}$, compared to the probability of failure $\frac{4}{E(Z_S)}$ which is inverse polynomial. The union bound, our only idea to handle these many-many overlapping k sets, would not give anything useful.

The problem is that Chebyshev's Inequality is too general for our purpose, it holds for *any random variable*. Our current random variable of interest Z_S is fortunately very special: it is the sum of INDEPENDENT indicator random variables and one can make use of this to improve the failure probability to inverse exponential.

Inequality 3 *(Chernoff's) Let Y be a binomial random variable of n independent trials of probability p. Then for every $a > 0$*

- $Pr(Y \leq np - a) \leq e^{-\frac{a^2}{2pn}}$
- $Pr(Y \geq np + a) \leq e^{-\frac{a^2}{2pn} + \frac{a^3}{2(pn)^2}}$
- *for every ϵ there is a $c = c(\epsilon)$ such that $Pr(|Y - E(Y)| \geq \epsilon E(Y)) \leq 2e^{-cE(Y)}$*

Proof. The trick is similar to the proof of Chebyshev, but instead of *squaring* our deviation of interest we *exponentiate* it, and then use Markov's Inequality. We cannot use symmetry when proving (i) and (ii), the difference between positive and negative deviation is real. Let $X = Y - np = \sum_{i=1}^{n} X_i$ where the X_i are independent, have expectation 0, and $Pr(X_i = 1 - p) = p$, $Pr(X_i = -p) = 1 - p$. Let $\lambda > 0$ be selected later.

(i) Then

$$Pr(X < -a) = Pr(e^{-\lambda X} > e^{\lambda a}) \leq \frac{E(e^{-\lambda X})}{e^{\lambda a}} = \frac{E(e^{-\lambda \sum X_i})}{e^{\lambda a}} = \frac{E(\prod_{i=1}^{n} e^{-\lambda X_i})}{e^{\lambda a}}.$$

Now the independence of the X_i comes into the picture: the expectation of the product is the product of expectations.

$$\begin{aligned} Pr(X < -a) &\leq \frac{\prod_{i=1}^{n} E(e^{-\lambda X_i})}{e^{\lambda a}} \\ &= \frac{e^{\lambda pn}(pe^{-\lambda}+1-p)^n}{e^{\lambda a}} \\ &\leq e^{\lambda pn - pn + pe^{-\lambda}n - \lambda a} \\ &= e^{pn(\lambda - 1 + e^{-\lambda}) - \lambda a} \\ &\leq e^{pn\frac{\lambda^2}{2} - \lambda a} \end{aligned}$$

Here the last inequality follows from the second order Taylor series estimate for $e^{-\lambda} < 1 - \lambda + \frac{\lambda^2}{2}$. The bound in (i) is then obtained by choosing $\lambda = \frac{a}{pn}$.

(ii) We start similarly.

$$Pr(X > a) = Pr(e^{\lambda X} > e^{\lambda a}) \leq \frac{E(e^{\lambda X})}{e^{\lambda a}} = \frac{\prod_{i=1}^{n} E(e^{\lambda X_i})}{e^{\lambda a}},$$

where the last equality follows again from the independence of X_i. Then

$$\begin{aligned} Pr(X > a) &= \frac{(pe^{\lambda(1-p)} + (1-p)e^{-\lambda p})^n}{e^{\lambda a}} \\ &= \frac{e^{-\lambda pn}(pe^{\lambda}+1-p)^n}{e^{\lambda a}} \\ &\leq e^{-\lambda a - pn(\lambda + 1 - e^{\lambda})} \end{aligned}$$

Substituting $\lambda = \ln(1+\frac{a}{pn})$ and using that then $\lambda > \frac{a}{pn} - \frac{1}{2}\left(\frac{a}{pn}\right)^2$ we get

$$\begin{aligned} Pr(X > a) &\leq e^{-(\frac{a}{pn} - \frac{a^2}{2(pn)^2})(a+pn) + pn(1-(1+\frac{a}{pn}))} \\ &\leq e^{-\frac{a^2}{pn} - a + \frac{a^3}{2(pn)^2} + \frac{a^2}{2pn} + a} \\ &= e^{-\frac{a^2}{2pn} + \frac{a^3}{2(pn)^2}} \end{aligned}$$

□

Remark The X_i do not have to have have the same distribution. The same Chernoff bounds hold if each Y is the union of n independent Bernoulli variable with probabilities $p_1, \dots, p_n$ and $pn := \sum p_i$. At some point in the proof one must use convexity and Jensen's Inequality.

Applying (i) for Z_S we have

$$Pr(S \text{ contains at most } \tfrac{1}{2}\tbinom{k}{2}p \text{ edges}) < e^{-\frac{E(Z_S)}{8}}.$$

Now we can use the union bound:

$$
\begin{aligned}
Pr(\exists S \in \binom{V}{k} \text{ containing at most } \tfrac{1}{2}\tbinom{k}{2}p \text{ edges}) &< \binom{n}{k} e^{-\frac{\binom{k}{2}p}{8}} \\
&\leq \exp(k(\ln n + 1 - \ln k - \frac{(k-1)p}{16})) \\
&\to 0,
\end{aligned}
$$

provided $k > \frac{16 \log n}{p} + 1$. Hence we intend to fix our target upper bound on the independence number to be

$$k = K\frac{\log n}{p} = \frac{K}{\epsilon}\sqrt{n}\log n,$$

where $K \geq 17$ is a large constant (to be determined later).

From now on we can think of all sets of size k to have at least

$$\frac{1}{2}\binom{k}{2}p \sim \frac{K^2}{4\epsilon}\sqrt{n}\log^2 n$$

edges. Our strategy requires the removal of the edges of a family of a maximal set of pairwise edge-disjoint triangles. The problem is if all the at least $\frac{1}{2}\binom{k}{2}p$ edges falling into some set S of size k are removed. Fix for the moment a set S of size k. We calculate the expected number W_S of triangles which have at least two vertices in S:

$$\mu = E(W_S) = (\binom{k}{2}(n-k) + \binom{k}{3})p^3 \approx \frac{k^2np^3}{2} = \frac{K^2\epsilon}{2}\sqrt{n}\log^2 n$$

This number is a good sign: there is a constant factor ϵ^2 less triangles are involved in S than the number of edges in S. We would be done if the the indicator variables of the triangles intersecting S in at least two vertices would be independent events. We could use Chernoff to get an inverse exponential bound on the failure probability (namely that two many triangles intersect S). But these triangle variables are of course not independent, so all we could use is Chebyshev, which would give us only an inverse-polynomial upper bound. And for the union bound we again need to sum up $\binom{n}{k}$ of them.

The following simple lemma gives a way out of this mess and uses the fact that we do not care if many triangles intersect S, we only care if many pairwise edge-disjoint ones do. It turns out that it is extremely unlikely that there is a family of 3μ pairwise edge-disjoint triangles that intersect S.

Lemma 1 *(Erdős-Tetali) Let $A_1, \ldots, A_m$ be events in an arbitrary probability space with $\mu = \sum_{i=1}^m Pr(A_i)$. Then for any s*

$$Pr(A_{i_1} \wedge \cdots \wedge A_{i_s} \textit{ holds for some } A_{i_1}, \ldots, A_{i_s} \textit{ mutually independent}) \leq \frac{\mu^s}{s!}$$

Proof. By the union bound the probability of our interest is at most

$$\begin{aligned}
\sum_{\substack{\{i_1,\ldots,i_s\} \\ A_{i_1},\cdots,A_{i_s}\text{independent}}} Pr(A_{i_1} \wedge \cdots \wedge A_{i_s}) &\leq \frac{1}{s!} \sum_{\substack{(i_1,\ldots,i_s) \\ A_{i_1},\cdots,A_{i_s}\text{independent}}} Pr(A_{i_1} \wedge \cdots \wedge A_{i_s}) \\
&= \frac{1}{s!} \sum_{\substack{(i_1,\ldots,i_s) \\ A_{i_1},\cdots,A_{i_s}\text{independent}}} Pr(A_{i_1}) \cdots Pr(A_{i_s}) \\
&\leq \frac{1}{s!} \sum_{i_1=1}^{m} \cdots \sum_{i_s=1}^{m} Pr(A_{i_1}) \cdots Pr(A_{i_s}) \\
&= \frac{1}{s!}(Pr(A_1) + \cdots + Pr(A_m))^s \\
&= \frac{\mu^s}{s!}
\end{aligned}$$

□

We use the lemma with $s = 3\mu$ for the indicator events of the triangles intersecting S in at least two vertices. We then have

$$Pr(\exists\ 3\mu \text{ pairwise edge-disjoint } K_3 \text{ intersecting } S \text{ in at least two points}) \leq \frac{\mu^s}{s!} < \left(\frac{e}{3}\right)^{3\mu}$$

So the probability that there is such a set of size k is upper bounded by the union bound:

$$\binom{n}{k}\left(\frac{e}{3}\right)^{3\mu} \leq \exp(k(\ln n + 1 - \ln k) - 3\mu\log(3/e)) \to 0,$$

provided $k \ln n < 3\mu \log(3/e) \approx \frac{3\log(3/e)}{2} k^2 np^3$. Substituting the values of k and p we obtain that we should select

$$K > \frac{2}{3\log(3/e)\epsilon^2}.$$

In conclusion, we proved that in $G(n, \epsilon n^{-1/2})$ a.a.s. every subset S of size k has the following two properties:

- the number of edges in S is at least $\frac{1}{2}\binom{k}{2}p$
- there are at most 3μ pairwise edge-disjoint triangles that intersect S in at least two vertices

Let H be an instance of such a graph. Remove the edges of an arbitrary maximal family of pairwise edge-disjoint triangles to obtain graph H'. Then H' is triangle-free and its independence number is at most $k = \frac{K}{\epsilon}\sqrt{n}\log n$. For this we need that the number of edges removed in each k-set is not more than the number of edges in the set. We remove at most 9μ edges in every k-set, so we just need that $\frac{9}{2}k^2np^3 \approx 9\mu < \frac{1}{2}\binom{k}{2}p \approx \frac{k^2p}{4}$, which holds if

$$\epsilon < \frac{1}{\sqrt{18}}.$$

This gives the following lower bound on $R(3,k)$, which is tight up to a factor of $\log^2 k$.

Theorem 8 *(Erdős, 1961)*

$$R(3,k) = \Omega\left(\frac{k^2}{\log^2 k}\right).$$

This lower bound on $R(3,k)$ was first proved by Erdős half a century ago by a very complicated probabilistic argument; the proof we gave above is due to Michael Krivelevich. There are a couple of alternative proofs that give similar bounds, all of them are non-constructive. The best explicit construction of a triangle-free graph has independence number $\Theta(n^{2/3})$ (due to Alon, using finite fields and codes)

5 Improving the upper bound on $R(3,k)$

We already saw that the $R(3,k)$ problem is an excellent terrain to demonstrate many of the standard probabilistic techniques (and we didn't even see everything). As far as history of the upper bounds goes, the quadratic bound on $R(3,k)$ coming from the Erdős-Szekeres proof was improved by Graver and Yackel (1966), by almost a full log-factor, to $O(\log\log k \frac{k^2}{\log k})$. The $\log\log$-factor waited till 1980 to be erased. This paper of Ajtai, Komlós and Szemerédi is a very important one in the development of the probabilistic combinatorics, as it is the origin of the *semi-random method* (also called *Rödl nibble method* after Rödl's ingenious proof of the Erdős-Hanani conjecture on almost perfect k-set covers in l-uniform hypergraphs). The next significant step in the $R(3,k)$ story was taken in 1995 when Jeong Han Kim, then a PhD-student of Jeff Kahn from Rutgers University, established that the true order of magnitude of $R(3,k)$ is $\frac{k^2}{\log k}$. The proof of correctness of his probabilistic construction is heavily based on martingales and is termed a technical tour de force. Kim received the Fulkerson prize for his achievement. Recently, Tom Bohman (curiously also a former student of Jeff Kahn) gave a new proof of the tight lower bound using a different technique. He managed to analyse the independence number of a well-studied graph process, the *triangle-free process*, using what is called the *differential equation method*. This method is the topic of our Doc-course.

In this section we prove the Ajtai-Komlós-Szemerédi Theorem. We will give a much simpler proof using nothing more than linearity of expectation. The argument is the modification by Alon of one of the proofs of Shearer and also gives a decent constant factor. However, we don't really try to optimize the constant factor, the best known upper bound is $(1 + o(1))\frac{k^2}{\log k}$ and is due to Shearer (1983).

The heart of the proof is a theorem improving Turán's Theorem for triangle-free graphs. By Theorem 4 we have that for *any* graph G, $\alpha(G) \geq \frac{n}{\bar{d}(G)+1}$, where $\bar{d}(G) = \frac{2e(G)}{n}$ is the average-degree of G. If G is *triangle-free*, then one can multiply the lower bound by an extra $\log \bar{d}(G)$-factor.

Theorem 9 *Let G be triangle-free with maximum degree $\Delta \geq 1$. Then*

$$\alpha(G) \geq \frac{n}{8\Delta} \log_2 \Delta.$$

Corollary 5 *Let G be triangle-free with average-degree $\bar{d} \geq 1$. Then*

$$\alpha(G) = \Omega\left(\frac{n}{\bar{d}} \log \bar{d}\right).$$

Proof of Corollary. Every graph G has a subgraph H on at least $n/2$ vertices with maximum degree at most $2\bar{d}(G)$. (Otherwise there would be more than $n/2$ vertices of degree larger than $2\bar{d}(G)$, implying $2e(G) > \frac{n}{2} \cdot 2\bar{d}(G) = 2e(G)$, a contradiction. Applying Theorem 9 in H we get

$$\alpha(G) \geq \alpha(H) \geq \frac{n}{16\bar{d}} \log_2(2\bar{d}).$$

□

Corollary 6 $R(3, k) < 8\frac{k^2}{\log_2 k}$

Proof of Corollary. Let G be a triangle-free graph on $n = 8\frac{k^2}{\log_2 k}$ vertices. We prove its independence number is at least k. If $\Delta(G) \geq k$ we are done, because the neighborhood of a vertex of maximum degree will be an independent set with order k.

If $\Delta(G) < k$, then we use Theorem 9 and get that

$$\alpha(G) \geq \frac{n}{8\Delta} \log_2 \Delta \geq \frac{8\frac{k^2}{\log_2 k}}{8k} \log_2 k = k.$$

□

Proof Theorem 9. We need to find a large independent set in G. The first ideas we had, like taking a random set by choosing every element with probability p or even the trickier proof idea of Theorem 4 will take us only so far: $\alpha(G) = O(\frac{n}{\Delta})$. Our problem in these approaches is that we do not really know how to use triangle-freeness to strengthen the bound.

One strange idea in the proof we present is to take a random independent set W, chosen uniformly at random among all independent sets of G and hope for it being large. Why, you could ask, do we want to take into account *all* independent sets, including singletons, or pairs of non-neighbor vertices, etc...? The answer is, as always, because we do not expect to lose much, but then we can analyse. Indeed, there are only n singletons, at most $\binom{n}{2}$ non-neighbors, but they will be overwhelmed by the many (independent) subsets of even just a single independent set I of size $k \approx \sqrt{n}$; the average size of independent sets contained in I is $(\sum_{i=0}^{k} i\binom{k}{i})2^k = \frac{k}{2}$, only a loss of constant factor 2.

The standard idea in order to calculate the expected size of W would be to introduce indicator random variables I_v for a vertex v being in W and represent $|W|$ as $\sum_v I_v$. The problem is that single vertices can behave very differently in terms of being in independent sets: no good uniform lower bound is possible. However, triangle-freeness implies that vertices *together with their neighborhood* exhibit a somewhat predictable behaviour: either the vertex is in the independent set and the neighbors are not or some of the neighbors are in there and the vertex is not. This motivates the following random variable for each $v \in V$

$$X_v = \Delta I_v + |N(v) \cap W|.$$

To estimate the expectation, let $U = U(v) = V \setminus (\{v\} \cup N(v))$ and let us look at the independent sets of G based on their intersection S with U.

Claim 1 *For every independent set $S \subseteq U$,*

$$E(X_v \mid S = W \cap U) \geq \frac{\log_2 \Delta}{4}.$$

Proof. For the given S, we define the set

$$Y = Y(S) = \{w \in N(v) : \ \{w\} \cup S \text{ is independent}\}$$

of eligible vertices that can in principle be in the same independent set with the whole of S. The crucial fact is that *any* subset Y' of Y can be added to S to form an independent set, since, G being triangle-free, the whole neighborhood $N(v)$ is independent. Let's say that $|Y| = y$. Then there are $2^y + 1$ ways to extend S into an independent set: either one can add v or some subset Y' of Y. Thus

$$E(X_v \mid S = W \cap U) = \frac{\Delta}{2^y + 1} + \frac{\sum_{i=0}^{y} i\binom{y}{i}}{2^y + 1} = \frac{\Delta + y2^{y-1}}{2^y + 1}.$$

This expression is at least $\frac{\log_2 \Delta}{4}$, otherwise

$$4\Delta - \log_2 \Delta < 2^y(\log_2 \Delta - 2y) < \sqrt{\Delta} \log_2 \Delta,$$

a contradiction for $\Delta \geq 16$. The last inequality follows because the left side is positive so $\Delta > 2^{2y}$. (For $\Delta < 16$ the whole statement of Theorem 9 is not very meaningful, one can check that Turán's bound for example gives the promised constant factor: $\alpha(G) \geq \frac{n}{\Delta+1} \geq \frac{n}{8\Delta}\log_2\Delta$.) $\square$

By the Claim the whole expectation $E(X_v)$ is at least $\frac{\log_2\Delta}{4}$, so for the sum $X = \sum_v X_v$ we have

$$E(X) \geq \frac{\log_2\Delta}{4}n.$$

Counting the contributions of the two main terms separately

$$X = \Delta\sum_{v\in V} I_v + \sum_{v\in V}|N(v)\cap W| \leq \Delta|W| + \Delta|W| = 2\Delta|W|.$$

Then for the expectation we have $E(2\Delta|W|) \geq E(X) \geq \frac{\log_2\Delta}{4}n$ and it follows that $E(|W|) \geq \frac{\log_2\Delta}{8\Delta}n$.

Hence *there exists* an independent independent set in G of the required size. $\square$

6 Graph processes

The topic of our course, the *differential equation method*, is particularly effective in analysing various graph processes. For example in the *triangle-free process* the edges of the complete graph come in a random order and are inserted unless they would close a triangle (in which case they are discarded). Then one tries to analyse various parameters of the graph obtained, say, its independent set.

In the following section first we introduce the most basic of all graph processes and chat about some of its properties. Then we focus on one of the properties important in further discussion: the creation of a *giant component*. We derive the classic theorem of Erdős and Rényi about the threshold probability when in $G(n,p)$ a component of linear size appears and we do this using another kind of process, the Galton-Watson stochastic process.

6.1 The random graph process

Let $e_1, \dots e_{\binom{n}{2}}$ be the edges of K_n. Choose a permutation $\pi \in S_{\binom{n}{2}}$ uniformly at random and define an increasing sequence of subgraphs (G_i) where $V(G_i) = V(K_n)$ and $E(G_i) = \{e_{\pi(1)}, \dots, e_{\pi(i)}\}$. It is clear that G_i is an n-vertex graph with i edges, selected uniformly at random from all n-vertex graphs with i edges. (BTW, $G(n,i)$ when $i \geq 0$ is an integer, denotes the *random graph with i edges*, an alternative random graph model to $G(n,p)$. The two models are highly related via the obvious parameter change and for most purposes one can easily switch between talking about $G(n,p)$ and $G(n, \lfloor p\binom{n}{2}\rfloor)$. Notice that ambiguity between the two random graph models can occur only if the second parameter is 0 or 1, but fortunately we rarely investigate these cases in any of the models.)

In the following we briefly summarize the most important properties of the random graph process in its early phases. All statements that follow are understood *asymptotically almost surely.*

For $i = 0$ the graph is empty, consists of n isolated vertices. Then independent edges start to creep in. Until $i \ll \sqrt{n}$ there is no degree 2 vertex, when $i \gg \sqrt{n}$ there are degree two vertices, when $i \gg n^{2/3}$ the first degree three vertex appears. For any constant k, when $n^{1-\frac{1}{k-1}} \ll i \ll n^{1-\frac{1}{k}}$ all trees on k already appear, but no trees on $k+1$ vertices do. Until $i \ll n$, the graph is still a forest with components of sub-logarithmic order $o_i(\log n)$.

The interesting phase is when $i = \frac{c}{2}n$ becomes linear in n (i.e., c is a constant). If $c < 1$ the largest component is of order $\Theta_c(\log n)$. At this point typically order $\log n$ sized components gobble up constant sized ones and grow pretty slowly. As c reaches 1, during a short critical period comparable sized components start to merge and hence the speed of growth of the size of the largest component becomes very fast. The larger the largest component becomes the more likely it is that it is him who eats up the smaller pieces (and not the smaller ones join). Within a small critical window of length about $\omega(n^{2/3})$ around $i = \frac{n}{2}$ the largest component becomes linear in n. In fact for $c > 1$ there is a single "giant component" of order $\Theta_c(n)$.

When i is as large as $\frac{\log n}{4}n$, then practically the whole graph is one big giant component and a few isolated vertices. The graph becomes connected when the last isolated vertex disappears, i.e., when $i = \frac{n}{2}(\log n + O(1))$, The same time the graph also has a perfect matching (provided n is even). Just a tiny bit later, when $i = \frac{n}{2}(\log n + \log\log n + O(1))$, the graph is 2-connected, Hamiltonian, and pancyclic even. To have minimum degree k or even k-connectivity for some constant k, we only need to increase the second term: it happens when $i = \frac{n}{2}(\log n + (k-1)\log\log n + O(1))$.

For these last properties a stronger statement is valid. Given a particular graph process (G_i) and a graph property $\mathcal{P}$ possessed by K_n, the *hitting time* $\tau(\mathcal{P}) = \tau(\mathcal{P}, (G_i))$ is the smallest i for which G_i has property $\mathcal{P}$. With this notation it is true that essentially the only reason the random graph is not connected is that it has an isolated vertex. The very moment when the last isolated vertex receives its first incident edge, the graph immediately becomes connected and also has a perfect matching. Formally,

$$\tau(\text{connected}) = \tau(\text{minimum degree is at least 1}) = \tau(\text{having perfect matching}).$$

Analogously, a trivial obstruction to Hamiltonicity and 2-connectivity is a vertex of degree less than 2. When the last vertex of degree one receives its second incident edge, via this very edge, the random graph becomes Hamiltonian and 2-connected.

$$\tau(\text{Hamiltonian}) = \tau(\text{minimum degree is at least 2}) = \tau(\text{2-connected})$$

Similarly, the bottleneck for k-connectivity is a vertex of degree $k-1$.

$$\tau(\text{minimum degree is at least } k) = \tau(k\text{-connected}).$$

6.2 The Galton-Watson process and the giant

The Galton-Watson process was introduced to model the alarming situation regarding the extinction of English aristocratic last names. What is the probability that a name becomes extinct?

In the simplified Galton-Watson model we assume that every individual has the same distribution X which describes the probabilities of the individual producing exactly i children. The process starts with a single node v_0 (Joel Spencer likes to call v_0 Eve). Eve produces children according to the distribution X. Then each of her children also produce children according to X, and these events are mutually independent from each other and also from the variable according to which Eve produced her children. And so on... Each children produces his children according to an independent instance of X until the end of time or until at some point every eligible descendant produces 0 children and the unfortunate event of the family name dying out occurs.

Formally speaking let us consider an infinite sequence of random variables $X_1, X_2, \ldots, X_i, \ldots$, each with the same distribution X and mutually independent. We start the process with a root called Eve, who is the first member of our population. In general we use the Breadth First Search traversing of the tree we created in order to determine who is the ith member of the population. During the process every member of the population has one shot to produce children: the ith member of the population produces children according to X_i. We call those members who haven't yet tried to produce children *alive*, and the ones who already had their shot we call *dead*.

The main question of English aristocrats is: what fertility distribution X should I and my male descendants have in order for our name to survive forever? The simple answer, given below in the basic theorem of Galton-Watson processes, depends only on the expectation: the name survives forever (essentially) if and only if $E(X) > 1$.

Formally, let the random variables Y_t (accounting for the number of alive vertices) be defined formally by the recursion

$$Y_0 \equiv 1, \qquad Y_t = Y_{t-1} + X_t - 1.$$

Indeed, in the tth round one of the Y_{t-1} live descendant produces its children via X_t and then dies. We say that the process dies out if $Y_t = 0$ for some t. The smallest t for which this happens is denoted by T. The *extinction probability* ρ_X of the distribution is

$$\rho_X = Pr(\exists t \ : \ Y_t = 0).$$

The following equation holds for $x = \rho_X$ because of the independence of the variables X_i.

$$x = \sum_{i=0}^{\infty} Pr(X = i)x^i =: f_X(x).$$

To explain: Eve produces i children with probability $Pr(X = i)$ and for each child the process has the same ρ_X probability of dying out. These are independent events for the children of Eve, so the probability of the process dying out conditioned on Eve having i children is ρ_X^i. Summing up for all i we get the extinction probability if the process started at Eve. The function $f_X(x)$ is the probability generating function of X.

Theorem 10 *(i)* $E(X) \le 1 \;\Rightarrow\; \rho_X = 1$ *unless* $Pr(X = 1) = 1$

(ii) $E(X) > 1 \;\Rightarrow\; \rho_X$ *is the unique solution of* $x = f_X(x)$ *in* $[0, 1)$

Examples.

1. Poisson distribution. Let P be a random variable of Poisson distribution $Po(c)$ with mean c. The probability generating function of P is

$$f_P(x) = \sum_{i=0}^{\infty} \frac{c^i}{i!} e^{-c} x^i = e^{c(x-1)}.$$

Later it will be convenient for us to use $\beta(c)$ for the *survival probability.*

$$1 - \beta(c) = \rho_P = e^{-\beta(c)c}.$$

To show (i) of Theorem 10 for the Poisson distribution one just needs to use Chernoff bounds (To have $Y_k > 0$ one needs $X_1 + \cdots + X_k \ge k$. Considering $E(X) = c < 1$, the expectation of $X_1 + \cdots + X_k$ is $ck < k$, so we would need to deviate from it by a positive fraction of the expectation; this is pretty (=exponentially) unlikely.) For (ii) we saw in the discussion before Theorem 10 that ρ_P must satisfy the given equation. One can then do some basic calculus and check that the function $f(y) = 1 - y - e^{-cy}$ has a single solution in $[0, 1)$. Finally, it was a HW exercise to exclude the possibility that the Poisson process dies out almost surely when $c > 1$.

2. Binomial distribution. Let Y_n be a random variable of binomial distribution $Bi(n, p)$ with mean $np \to c$ is a constant. (We suppress the reference to p, since this will not change in our discussion) Then

$$f_{Y_n}(x) = \sum_{i=1}^{n} \binom{n}{i} p^i (1-p)^{n-i} x^i = (1 - p + xp)^n \approx e^{-pn(1-x)}$$

and hence for every $x \in \mathbb{R}$ we have that

$$\lim_{n\to\infty} f_{Y_n}(x) = e^{c(x-1)} = f_P(x).$$

In particular the extinction probability $\rho_{Y_n} =: \rho(n)$ of the process defined by $Bi(n,p)$ converges to $\rho_P = 1 - \beta(c)$, the extinction probability defined by $Po(c)$.

Our goal is to analyse the component structure of the the random graph $G(n,p)$ for the interesting sparse range of p, i.e., when p is $\frac{c}{n}$ for some constant c. We pick a vertex and generate a spanning tree in its component by mimicking a Galton-Watson-type process. First we let v_1 randomly select its neighbors according to a random variable X_1 with distribution $Bi(n-1,p)$. Suppose the children are $v_2, v_3, \ldots, v_{X_1+1}$. Then we let v_2 select its neighbors according to a random variable X_2 with distribution $Bi(n - X_1 - 1, p)$ from the vertices in $V \setminus \{v_1, \ldots, v_{X_1+1}\}$ (we do *not* care whether v_2 has edges to the other neighbors of v_1). Then we let v_3 select its neighbors according to a distribution $Bi(n - X_2 - X_1 - 1, p)$. And so on... We proceed along a BFS of the tree we create and for the ith vertex we select as many children from the still isolated ones as many the random variable X_i with distribution $Bi(n - \sum_{j=0}^{i} X_j, p)$ tells us.

This process described above is *not* a Galton-Watson process, because

1. the distributions we sample depend on the history of the process
2. in particular the distributions are *not* the same for every vertex.

Both of these concerns are forgotten if we are willing to see the world with a normal eye but through -3.00 dioptre glasses. The difference between $Bi(n - \sum_{j=0}^{i} X_j, p)$ and $Bi(n,p)$ then blurs away provided $\sum_{j=0}^{i} X_j$ is smaller order than n.

Theorem 11 *(Erdős-Rényi) Let $p = \frac{c}{n}$, where $c > 0$ is a constant.*

(i) If $c < 1$, then a.a.s. the largest component of $G(n,p)$ has at most $\frac{4}{(1-c)^2}\log n$ vertices

(ii) If $c > 1$, then a.a.s

- *$G(n,p)$ contains a single giant component of $(1+o(1))\beta n$ vertices*
- *the second largest component has at most $\frac{16c}{(c-1)^2}\log n$ vertices.*

Proof. (Janson-Łuczak-Ruciński ??)

(*i*) Let $pn = c < 1$ be given and let $C(v)$ denote the component of vertex v in $G(n,p)$. Let $k = \frac{4}{(1-c)^2}\log n$ be the threshold given in the statement for the largest component size.

Let us fix a vertex v and create a spanning tree of its component with our branching process. If it so happens that $C(v)$ contains more than k vertices, then during the first k sampling altogether the process must have created at least a combined k children for the first k members of the population. Hence

$$Pr(|C(v)| > k) \le Pr(\sum_{i=1}^{k} X_i \ge k) \le Pr(\sum_{i=1}^{k} X_i^+ \ge k).$$

Here on the right hand side each of the variables X_i^+ are binomially distributed $Bi(n,p)$. The second inequality is valid since each of the first k variables X_i have distributions of the form $Bi(n-m_i,p)$ for some $m_i \geq 1$, so we only increase the chance of their sum reaching k if we perform more trials. The sum $\sum_{i=1}^k X_i^+$ has distribution $Bi(kn,p)$ with mean $knp = ck < k$, so we can apply Chernoff with deviation $(1-c)k$:

$$Pr(\sum_{i=1}^{k} X_i^+ \geq k) \leq e^{-\frac{(1-c)^2k^2}{2ck}+\frac{(1-c)^3}{2(ck)^2}} = e^{-(\frac{(1-c)^2}{2c}\cdot\frac{2c-1}{c})k+O(1)} < e^{-\frac{(1-c)^2}{3c}k+O(1)} = O(n^{-4/3}).$$

In the last inequality we assumed $c > \frac{3}{4}$ (which is not a problem since we are proving a monotone statement).

(ii) Assume now that $np = c > 1$ and let $k_- = \frac{16c}{(c-1)^2}\log n$ and $k_+ = n^{2/3}$.

We proceed in three steps. We show first that a.a.s. for every vertex v and every $k, k_- \leq k \leq k_+$ we have that either the process started at v terminated within k_- steps or after the kth round there are at least $\frac{c-1}{2}k$ live vertices.

Fix a v and a k and let $Fail_{v,k}$ be the event that the process started at v did not terminate within k steps and after the kth round there are less than $\frac{c-1}{2}k$ live vertices. If this happens than the first k random variables X_i produced less than $k + \frac{c-1}{2}k = \frac{c+1}{2}k$ children. In particular each X_i has a binomial distribution $Bi(n-m_i,p)$ with $m_i \leq \frac{c+1}{2}k$. So

$$Pr(Fail_{v,k}) \leq Pr\left(\sum_{i=1}^{k} X_i \leq \frac{c+1}{2}k\right) \leq Pr\left(\sum_{i=1}^{k} X_i^- \leq \frac{c+1}{2}k\right),$$

where each X_- is a random variable with distribution $Bi(n-\frac{c+1}{2}k,p)$. The expectation of $\sum_{i=1}^k X_i^-$ is $k(n-\frac{c+1}{2}k)p \approx knp = ck > k$, so we have a chance of applying Chernoff bounds with deviation $\approx (c-\frac{c+1}{2})k = \frac{c-1}{2}k$ and obtain that

$$Pr\left(\sum_{i=1}^{k} X_i^- \leq \frac{c+1}{2}k\right) < e^{-\frac{(c-1)^2k^2}{8ck}(1+o(1))} < e^{-\frac{(c-1)^2}{9c}k} = n^{-16/9}$$

for large n. By the union bound

$$Pr\left(\bigvee_{v\in V}\bigvee_{k=k_-}^{k^+} Fail_{v,k}\right) < n\cdot n^{2/3}\cdot n^{-16/9} \to 0.$$

Hence from now on we assume (since it is true a.a.s.) that for every vertex v the component of v has at most k_- vertices, in which case we will call v *small vertex*, or the component of v has at least k_+ vertices, in which case the v is called a *large vertex*. Moreover, we also have that a.a.s. our process started at a large vertex v has at least $\frac{c-1}{2}k_+$ live vertices after the k_+th round.

Secondly, we prove that a.a.s. there is a single component which contains all large vertices. Let v' and v'' be two large vertices. First grow the process from v'. Then either we reach v'' and we are done. Otherwise we stop after k_+ rounds and identify a set $W_{v'}$ of $\frac{c-1}{2}k_+$ live vertices in $C(v')$. Then we grow the process from v''. Again, either we reach v' and we are done or we stop after k_+ rounds and identify a set $W_{v''}$ of $\frac{c-1}{2}k_+$ live vertices in $C(v'')$. The edges between $W_{v'}$ and $W_{v''}$ were not yet explored and the presence of any one of these edges would make v' and v'' be in the same component. The probability of this not happening is

$$(1-p)^{\frac{(c-1)^2}{4}k_+^2} < \exp(-\frac{c}{n}\frac{(c-1)^2}{4}n^{4/3}) = \exp(-\Theta(n^{1/3}))$$

So by the union bound any two large vertices are in the same component with probability $1 - n^2\exp(-\Theta(n^{1/3})) \to 1$.

Finally we need to bound the size of the component containing all large vertices. First we estimate the probability of a vertex being small. Recall that $\rho(n)$ is the extinction probability of the Galton-Watson process with the binomial distribution $Bi(n,p)$. We claim that for the probability $Pr(v \text{ is small}) := \nu(n,p)$ that v is small we have

$$\rho(n) - o(1) \leq \nu(n,p) \leq \rho(n-k_-).$$

For the second inequality note that if v is small than the process growing from it will encounter binomial random variables $Bi(n-m_i)$ with $m_i \leq k_-$. For the first inequality we compare our process to the Galton-Watson with $Bi(n,p)$: this process makes more trials at every step, the probability of its extinction minus the probability that it dies after k_- steps is the lower bound. (The probability that the Galton-Watson with $Bi(n,p)$ process dies after k_- steps is at most $Pr(X_1 + \cdots + X_k \leq k)$ for some $k \geq k_-$ where each X_i has distribution $Bi(n,p)$. This probability is at most $e^{\frac{(c-1)^2}{c}k}$ by the Chernoff bound and even summing them up for all $k \geq k_-$ it tends to 0.)

When n tends to infinity, then $\rho(n)$ and $\rho(n-k_-)$ both tend to the extinction probability $\rho_P = 1 - \beta(c)$ of the Poisson process with mean c. Hence for the number Y of small vertices we have that its expectation is $(1-\beta(c)+o(1))n$. In order to complete the proof of (ii) we still need to show Y is concentrated about its mean, then of course the size of the largest component is a.a.s. $\beta(c)n$. So we calculate the variance. Let I_v be the indicator variable of the event that v is small. Let us see first what is the probability that some $w \neq v$ is small assuming that v is small:

$$Pr(I_w = 1 | I_v = 1) \leq \frac{k_-}{n} + \nu(n-k_-, p)$$

To see this grow first the spanning tree of the (small) component of v. The first term accounts for the probability that w is in $C(v)$. For the second term grow the component of w on the vertices of $V \setminus \{v\} \cup N(v)$ and note that $\nu(r',p) \geq \nu(r'',p)$

for $r' \leq r''$. Then for the variance we have

$$\begin{aligned}
Var(Y) &= \sum_v \sum_w (E(I_v I_w) - E(I_v)E(I_w)) \\
&= E(Y) + \sum_v Pr(I_v = 1) \sum_{w \neq v} Pr(I_w = 1 | I_v = 1) - E(Y)^2 \\
&\leq E(Y) + \sum_v Pr(I_v = 1)(n-1)(\frac{k_-}{n} + \nu(n - k_-, p)) - E(Y)^2 \\
&\leq E(Y) + E(Y)(k_- + n\nu(n - k_-, p)) - E(Y)^2 \\
&= o(E(Y)^2)
\end{aligned}$$

For the last equality we note that $\nu(n - k_-, p)$ tends to $1 - \beta(c)$ and $k_- = o(n)$, so $k_- + n\nu(n - k_-, p) = (1 + o(1))E(Y)$. Chebyshev's Inequality then gives the concentration about the mean and (ii) follows. $\square$

References

[1] N. Alon, J. Spencer, **The Probabilistic Method**, 3rd ed. 2008, John Wiley.

[2] S. Janson, T. Łuczak, A. Ruciński, **Random graphs**, 2000, John Wiley.