Random Evolution of Graphs

Matija Bucić

October 15, 2014

0 Introduction

In this essay we present some famous results about the threshold functions for different properties in the Erdős-Rényi model and their relations to hitting times in the random evolution processes. In the end we present and prove the famous theorem about the sudden emergence of the giant component.

The results and proofs in this paper are based on the ones given by Bollobás[1], Komlós and Szemerédi[2] and Janson, Łuczak and Ruciński[3].

We will work with two random graph models:

- The Erdős-Rényi(ER) model $G(n, p)$ assigns a probability distribution on the set of all graphs with n vertices in a way that each edge exists with probability p and does not with probability $1 - p$.

  Therefore, a graph with M edges has probability $p^M(1-p)^{N-M}$ of being chosen where $N = \frac{n(n-1)}{2}$. .

- The uniform model $G(n, m)$ assigns a uniform distribution on the graphs with exactly m edges.

  As there are exactly $\binom{N}{m}$ graphs on m edges in this case, the probability of a specific one being chosen is $\frac{1}{\binom{N}{m}}$.

Throughout this essay, unless otherwise stated n will denote the number of vertices, $N = \frac{n(n-1)}{2}$ is the total number of possible edges, p will denote the edge probability and m or M will denote the number of edges in the uniform model.

We will often use the Stirling's formula to approximate factorials, the following is the strongest version we use:

$$\sqrt{2\pi s}\left(\frac{s}{e}\right)^s \leq s! \leq e\sqrt{s}\left(\frac{s}{e}\right)^s$$

Throughout the essay we use statements like: almost surely, with high probability all meaning that probability tends to 1 as $n \to \infty$.

1 The Clique Number

The aim of this section is to introduce some basic ideas we will often use afterwards, as well as to give an example as how peculiar the ER-model can actually be.

Throughout this section only p will be fixed and will not depend on n.

The clique number of a graph G is the maximal order of a complete subgraph and in this section we investigate it in the ER-model.

We start by defining the random variable X_d to be the number of complete subgraphs of order d in the graph $G(n,p)$.

The clique number is $d-1$ for the smallest d such that $X_d = 0$. This follows as if we have a $d-$complete subgraph by removing any vertex we get a $d-1-$complete graph.

We start by finding $\mathbb{E}(X_d)$. For this purpose let us define the indicator random variables Y_{G_i} for each subgraph on d vertices G_i, taking values 1 and 0 depending on whether G_i is complete or not, respectively. We also note that there are $\binom{n}{d}$ subgraphs of order d.

We have:

$$X_d = \sum_{i=1}^{\binom{n}{d}} Y_{G_i}$$

.

Taking expectation and using the standard result on additivity of expectation we get:

$$\mathbb{E}(X_d) = \mathbb{E}\left(\sum_{i=1}^{\binom{n}{d}} Y_{G_i}\right) = \sum_{i=1}^{\binom{n}{d}} \mathbb{E}(Y_{G_i}) = \sum_{i=1}^{\binom{n}{d}} \mathbb{P}(Y_{G_i} = 1)$$

Where the last equality follows by $\mathbb{E}(Y_{G_i}) = 0 \cdot \mathbb{P}(Y_{G_i} = 0) + 1 \cdot \mathbb{P}(Y_{G_i} = 1) = \mathbb{P}(Y_{G_i} = 1)$

Finally, we have $\mathbb{P}(Y_{G_i} = 1) = \mathbb{P}(G_i \text{ is complete}) = p^{\frac{d(d-1)}{2}}$.
So the desired expectation is:

$$\mathbb{E}(X_d) = \binom{n}{d} p^{\frac{d(d-1)}{2}}$$

We now turn to the main result of this section.

Let $d = d(n)$ be the greatest integer such that

$$\mathbb{E}(X_d) = \binom{n}{d} p^{\frac{d(d-1)}{2}} \geq \log n$$

We note that this is well defined as $\mathbb{E}(X_1) = n$ and $\mathbb{E}(X_n) = p^{\frac{d(d-1)}{2}} < 1$.

Theorem 1. Let $0 < p < 1$ be fixed. Then the clique number of $G(n,p)$ is either d or $d+1$ with probability tending to 1 as n tends to infinity.

The really remarkable thing about this theorem is that the window is so small. Intuitively this is due to the fact that once we have relatively few $d+1-$complete subgraphs, we consequently have even fewer $d+2-$complete subgraphs and each $d+2-$complete subgraph contains $d+2$ distinct $d+1-$complete subgraphs, while there are relatively few $d+2-$complete subgraphs sharing some $d+1-$complete subgraph.

The proof uses 2 ideas which we will meet again and again through this essay.
We need:

- To show $\mathbb{P}(X_{d+2} > 0) \to 0$ we will use $\mathbb{P}(X > 0) < \mathbb{E}(X)$ so if $E(X) \to 0$ so does $\mathbb{P}(X > 0)$.
- To show $\mathbb{P}(X_d > 0) \to 1$ we will show $\mathbb{E}(X) \to \infty$ and show that the variance is small enough for this to imply $\mathbb{P}(X_d = 0) \to 0$. We do this formally by using the Chebyshev's Inequality to show $\mathbb{P}(X = 0) \leq \frac{\mathrm{Var}(X)}{\mathbb{E}(x)^2}$.

Proof.

- Showing $\mathbb{P}(X_{d+2} > 0) \to 0$

$$\log n > \binom{n}{d+1} p^{\frac{d(d+1)}{2}} = \frac{n \cdot (n-1) \cdots (n-d)}{(d+1)!} p^{\frac{d(d+1)}{2}} > \frac{n \cdot (n-d)^d}{(d+1)!} p^{\frac{d(d+1)}{2}}$$

We now use $n > \log n$ and Stirling's Inequality $(d+1)! \leq e\sqrt{d+1}\left(\frac{d+1}{e}\right)^{d+1}$ to get:

$$1 > \frac{n}{\log n} \frac{(n-d)^d}{(d+1)!} p^{\frac{d(d+1)}{2}} \geq \frac{(n-d)^d e^d}{\sqrt{d+1}(d+1)^{d+1}} p^{\frac{d(d+1)}{2}}$$

$$(d+1)^{\frac{3}{2}} (d+1)^d > e^d (n-d)^d p^{\frac{d(d+1)}{2}}$$

$$(d+1)^{\frac{3}{2d}} (d+1) > e(n-d) p^{\frac{(d+1)}{2}}$$

Now we notice from the Taylor expansion $d+1 < 1 + \frac{2}{3}d + \frac{2}{9}d^2 < e^{\frac{2d}{3}}$ for any $d \geq 2$ so:

$$p^{\frac{(d+1)}{2}} < \frac{d+1}{n-d} \quad (1)$$

Now $\mathbb{P}(X_{d+2} > 0) \leq \mathbb{E}(X_{d+2}) = \frac{n-d-1}{d+2} p^{d+1}\, \mathbb{E}(X_{d+1}) < \frac{n-d}{d+1} p^{d+1} \log n < \frac{d+1}{n-d} \log n$.

Let $t = \lceil 4 \log_b n \rceil$, where $b = \frac{1}{p}$.

$$\binom{n}{t}p^{\frac{t(t-1)}{2}} \le \binom{n}{t}p^{2\log_b(n)(t-1)} \le n^t n^{-2(t-1)} = n^{2-t} < \log n$$

For large enough n. Hence, by definition of d we have $d \le t$.

Now we deduce there is a constant c such that for large enough n:

$$\mathbb{P}(X_{d+2} > 0) < \frac{d+1}{n-d}\log n < \frac{c(\log n)^2}{n} \to 0$$

.

- $\mathbb{P}(X_d = 0) \to 0$
  Let $\mu = \mathbb{E}(X_d)$.

$$\mathbb{P}(X_d = 0) \le \mathbb{P}(|X_d - \mu| \le \mu) \le \frac{\mathbb{E}(|X_d - \mu|^2)}{\mu^2} = \frac{\mathrm{Var}(X_d)}{\mu^2}$$

Where the second inequality follows from The Chebyshev Inequality.

We have $\mathrm{Var}(X_d) = \mathbb{E}(X_d^2) - \mu^2$, so now we find $\mathbb{E}(X_d^2)$.
Using the same notation as before and summing up over pairs of d-complete subgraphs having l vertices in common we get:

$$\mathbb{E}(X_d^2) = \mathbb{E}\left(\sum_{i,j=1}^{\binom{n}{d}} Y_{G_i} \cdot Y_{G_j}\right) = \sum_{i,j=1}^{\binom{n}{d}} 1 \cdot \mathbb{P}(Y_{G_i} = 1, Y_{G_j} = 1) =$$

$$\sum_{l=0}^{d}\sum_{i,j=1}^{\binom{n}{d}} \mathbb{P}(G_i, G_j \text{ are complete and share } l \text{ vertices.}) =$$

$$\sum_{l=0}^{d}\binom{n}{d}\binom{d}{l}\binom{n-d}{d-l}p^{2\binom{d}{2}-\binom{l}{2}}$$

We also have:

$$\mu^2 = \binom{n}{d}^2 p^{2\binom{d}{2}} = \binom{n}{d}p^{2\binom{d}{2}}\sum_{l=0}^{d}\binom{d}{l}\binom{n-d}{d-l}$$

So by combining these results we get:

$$\frac{\mathrm{Var}(X_d)}{\mu^2} = \frac{\mathbb{E}(X_d^2) - \mu^2}{\mu^2} = \sum_{l=0}^{d}\binom{n}{d}^{-1}\binom{d}{l}\binom{n-d}{d-l}\left(p^{-\binom{l}{2}} - 1\right)$$

Cancelling out the $l = 0, 1$ terms and ignoring the leftover contribution of $-\mu^2$ we get:

$$\frac{\mathrm{Var}(X_d)}{\mu^2} \le \sum_{l=2}^{d} \binom{n}{d}^{-1} \binom{d}{l} \binom{n-d}{d-l} p^{-\binom{l}{2}} \le 2 \sum_{l=2}^{d-1} \frac{d!^2}{l!(d-l)!^2} n^{-l} p^{-\binom{l}{2}} + \frac{1}{\mu} = 2 \sum_{l=2}^{d-1} \epsilon_l + \frac{1}{\mu}$$

Where we have defined $\epsilon_l = \frac{d!^2}{l!(d-l)!^2} n^{-l} p^{-\binom{l}{2}}$ and used $\frac{(n-d)!^2}{n!(n-2d+l)!} \le 2n^{-l}$ which we will now prove.

$$\frac{(n-d)!^2}{n!(n-2d+l)!} n^l = \frac{(n-d)(n-d-1)\cdots(n-2d+l+1)n^l}{n(n-1)\cdots(n-d+1)} \le$$

$$\le \frac{n^l}{n(n-1)\cdots(n-l+1)} \le \frac{n^d}{n(n-1)\cdots(n-d+1)} \le \left(\frac{n}{(n-d+1)}\right)^d =$$

$$= \left(1 + \frac{d-1}{(n-d+1)}\right)^d \le e^{\frac{d(d-1)}{n-d+1}} \to 1$$

Proposition 1.1. ϵ_l as defined above are first decreasing and then increasing on the interval $3 \le l \le d-1$, assuming n is large enough.

Proof. We have:

$$\frac{\epsilon_{l+1}}{\epsilon_l} = \frac{(d-l)^2}{(l+1)n} p^{-l}$$

if $l \le c \log_b n$ for some constant $c < 1$ then:

$$\frac{\epsilon_{l+1}}{\epsilon_l} \le \frac{(d-l)^2}{(l+1)} n^{c-1} \to 0 \text{ as } n \to \infty$$

as $l \le d = O(\log n)$

if $l \ge c \log_b n$ for some constant $c > 1$ then:

$$\frac{\epsilon_{l+1}}{\epsilon_l} \le \frac{(d-l)^2}{(l+1)} n^{c-1} \to \infty \text{ as } n \to \infty$$

Hence, we conclude that proposition holds unless $0.99 \log_b n \le l \le 1.01 \log_b n$.

Now consider the function $f = \frac{(d-x)^2}{(x+1)n} p^{-x}$ on the interval $I = [0.99 \log_b n, 1.01 \log_b n]$.

Differentiating we get $f' \ge 0 \Leftrightarrow (d-x)\left(\log b - \frac{1}{x+1}\right) \ge 2$ and we do have that if n is large enough this always holds, as $d - x \ge \frac{1}{2} \log_b n \to \infty$ and $\frac{1}{x+1} \le \frac{1}{0.99 \log_b n} \to 0$.

So the ratio of consecutive ϵ_l on this interval is increasing, which implies that ϵ_l is first decreasing and then increasing

$\square$.

An immediate corollary is $\epsilon_l \leq \epsilon_3 + \epsilon_{d-1}$.

So now we can bound:

$$\frac{\mathrm{Var}(X_d)}{\mu^2} \leq 2\left(\sum_{l=2}^{d-1} \epsilon_{d-1}\right) + \frac{1}{\mu} \leq 2\epsilon_2 + 2d(\epsilon_3 + \epsilon_{d-1}) + \frac{1}{\mu}$$

We now notice that $\mu \geq \log n \to \infty$ and using (1) we get:

$$d\epsilon_{d-1} = d!d^2 n^{1-d} p^{-\binom{d-1}{2}} = d^2 n p^{d-1} \epsilon_d \leq \frac{d^2(d+1)^2 d}{(n-d)^2 \log n} p^{-2} \to 0$$

$$d\epsilon_3 = d\frac{d!^2}{6(d-3)!^2} n^{-3} p^{-3} \leq d^7 n^{-3} p^{-3} \to 0$$

$$\epsilon_2 = \frac{d!^2}{2(d-2)!^2} n^{-2} p^{-1} \leq d^4 n^{-2} p^{-1} \to 0$$

So we have proved

$$\frac{\mathrm{Var}(X_d)}{\mu^2} \to 0$$

and therefore $\mathbb{P}(X_d = 0) \to 0$, as desired.

$\square$

2 Connectivity

In this section we find the best threshold functions for the property of being connected, introduce the idea of random evolution of a graph and its relation to the threshold functions through hitting times.

From now onwards $p = p(n)$ is allowed to vary with n.

We define $p_u(n)$ to be an upper threshold function for the property Q if and only if $\mathbb{P}(G(n, p_u(n))$ has $Q) \to 1$ as $n \to \infty$, while $p_l(n)$ is a lower threshold function for the property Q if and only if $\mathbb{P}(G(n, p_l(n))$ has $Q) \to 0$ as $n \to \infty$.

Theorem 2. Let $\omega(n) \to \infty$, then $p_l = \frac{\log n - \omega(n)}{n}$, $p_u = \frac{\log n + \omega(n)}{n}$ are lower and upper threshold functions for the property of being connected, respectively.

Intuitively, this theorem tells us that there are certain boundaries such that if probability for choosing an edge is smaller, we almost certainly do not have the property, while if probability is larger, it is almost certain that we do have the property. Once again, the size of the uncertainty window is quite remarkable.

Proof. We start by showing that we may assume $\omega(n)$ is growing arbitrarily slowly.

Definition. We say a property Q of graphs is increasing if G having Q implies that for any $e \in E(G)$ graph $G + e$ also has Q.

Lemma 2.1. Let Q be an increasing property, given $p_1 \leq p_2$ then

$$\mathbb{P}(G(n, p_1) \text{ has } Q) \leq \mathbb{P}(G(n, p_2) \text{ has } Q)$$

Proof. We pick an element $G_1 \in G(n, p_1)$ and $G \in G(n, p)$, where p remains to be found. We notice that for $G_2 = G_1 \cup G$ each edge is chosen with probability $p_1 + (1 - p_1)p$. Furthermore, if G_1 has Q then so does G_2, as Q is increasing. Hence, we want to pick p such that $p_2 = p_1 + (1 - p_1)p$ so we choose $p = \frac{p_2 - p_1}{1 - p_1}$ and this implies $\mathbb{P}(G_1$ has $Q) \leq \mathbb{P}(G_2$ has $Q)$.

□

Hence, we may assume $\omega(n) < \log \log \log n$, because if the results is true for all such ω, it is true for any larger as well by **Lemma** 2.1.

We will assume that n is large enough to make $\omega > 10$.

Let us define X_k to be the number of components of $G \in G(n, p)$ having exactly k vertices.

- We show $p = p_l = \frac{\log n - \omega(n)}{n}$ is a lower threshold function.

  We apply the same method as in the second part of **Theorem** 1 to show $\mathbb{P}(X_1 = 0) \to 0$ implying that there is almost always an isolated vertex and therefore, it is almost always disconnected.

As before, we need to find $\mathbb{E}(X_1)$ and $\mathbb{E}(X_1^2)$.

Let Y_i be the indicator function of vertex i being isolated.

We use $X_1 = \sum_{i=1}^n Y_i$ to get:

$$\mu = \mathbb{E}(X_1) = \sum_{i=1}^n \mathbb{E}(Y_i) = \sum_{i=1}^n \mathbb{P}(Y_i = 1) = n(1-p)^{n-1} \leq ne^{-p(n-1)} =$$

$$= ne^p e^{-\log n + \omega(n)} = e^{p+\omega n} \to \infty$$

Where we have used $1 - p \leq e^{-p}$ and the fact that $\omega(n) \to \infty$.

$$\mathbb{E}\left(X_1^2\right) = \mathbb{E}\left(\left(\sum_{i=1}^n Y_i\right)^2\right) = \sum_{i=1}^n \mathbb{P}\left(Y_i^2 = 1\right) + \sum_{i \neq j}^n \mathbb{E}(Y_i Y_j) =$$

$$= \sum_{i=1}^n \mathbb{P}(Y_i = 1) + \sum_{i \neq j}^n \mathbb{P}(Y_i, Y_j = 1)$$

$$\mathbb{E}\left(X_1^2\right) = \mu + n(n-1)\mathbb{P}(Y_1, Y_2 = 1) = \mu + n(n-1)(1-p)^{2n-3}$$

Where the last equality follows as in order for both $1, 2$ to be isolated, we must not have chosen $2n - 3$ edges containing any of them.

Using the same inequality as before, we get:

$$\mathbb{P}(X_1 = 0) \leq \frac{\mathrm{Var}(X_d)}{\mu^2} = \frac{\mathbb{E}\left(X_d^2\right) - \mu^2}{\mu^2} = \frac{\mu + n(n-1)(1-p)^{2n-3} - n^2(1-p)^{2n-2}}{\mu^2} =$$

$$\frac{1}{\mu} + \frac{np - 1}{n(1-p)} = \frac{1}{\mu} + \frac{\log n - \omega(n) - 1}{n - \log n + \omega(n)} \leq \frac{1}{\mu} + \frac{2\log n}{n} \to 0$$

Where we have used that for large enough n, we have $n - \log n + \omega(n) \geq \frac{n}{2}$, $\frac{2\log n}{n} \to 0$ and $\mu \to \infty$.

This shows p_l is a lower threshold function for the property of being connected.

- We show $p = p_u = \frac{\log n + \omega(n)}{n}$ is an upper threshold function.

For this part we will use a similar idea to the one used in the first part of **Theorem** 1.

We start with a simple observation that G is disconnected if and only if there is a component of size at most $\left\lfloor \frac{n}{2} \right\rfloor$. Using this, we get:

$$\mathbb{P}(G(n,p) \text{ is disconnected}) = \mathbb{P}\left(\sum_{i=1}^{\lfloor \frac{n}{2} \rfloor} X_i \geq 1\right) \leq \mathbb{E}\left(\sum_{i=1}^{\lfloor \frac{n}{2} \rfloor} X_i\right) = \sum_{i=1}^{\lfloor \frac{n}{2} \rfloor} \mathbb{E}(X_i)$$

Where the first equality is due to the observation above, the second inequality holds because $\mathbb{E}(X) = \sum_{i=0}^{\infty} i\mathbb{P}(X = i) \geq \sum_{i=1}^{\infty} \mathbb{P}(X = i) = \mathbb{P}(X \geq 1)$, while the third equality is due to the additivity of expectation.

We now estimate $\mathbb{E}(X_i)$.

Let Y_i be the indicator function for the i-th k-subset of vertices of n then $X_i = \sum_{i=1}^{\binom{n}{k}} Y_i$

$$\mathbb{E}(X_i) = \mathbb{E}\left(\sum_{i=1}^{\binom{n}{k}} Y_i\right) = \sum_{i=1}^{\binom{n}{k}} \mathbb{E}(Y_i) = \sum_{i=1}^{\binom{n}{k}} \mathbb{P}(Y_i = 1) = \binom{n}{k}\mathbb{P}(Y_1 = 1)$$

If we denote by $q_k = \mathbb{P}(G(k,p) \text{ is connected})$ we have $\mathbb{P}(Y_1 = 1) = (1-p)^{k(n-k)}q_k$, as we have to guarantee that there are no edges between our k chosen vertices and the remainder of $n-k$ vertices. Until now we have just had equalities, but now it seems rather hard to find q_k explicitly. It turns out that $q_k \leq 1$ is enough. We may note that this does not weaken our expression by much, as **Theorem** 2 itself implies $q_k \to 1$ as $k \to \infty$. Using this, we now get:

$$\mathbb{P}(G(n,p) \text{ is disconnected}) = \sum_{i=1}^{\lfloor \frac{n}{2} \rfloor} \mathbb{E}(X_i) \leq \sum_{i=1}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{k}(1-p)^{k(n-k)}$$

We now have to argue differently depending on whether k is small or large.

- If $k \leq n^{\frac{3}{4}}$ we have:

  $$\binom{n}{k}(1-p)^{k(n-k)} \leq \frac{n(n-1)\cdots(n-k+1)}{k!}e^{-pk(n-k)} \leq$$
  $$\frac{n^k}{\left(\sqrt{2k\pi}\frac{k}{e}\right)^k}e^{-pk(n-k)} \leq \left(\frac{ne}{k}\right)^k e^{-pk(n-k)}$$

  Where we have first used the inequality $1-p \leq e^{-p}$ and then Stirling's Inequality.

  $$\left(\frac{ne}{k}\right)^k e^{-pk(n-k)} = e^{k(\log n + 1 - \log k - pn + pk)} =$$
  $$e^{k(1+\log n - \log k - \log n - \omega(n) + pk)} \leq e^{k(1-\omega(n)+1)} = e^{-k(\omega(n)-2)}$$

  Where we have $pk \leq (\log n + \omega(n))n^{-\frac{1}{4}} \to 0$. So for large enough n we have $pk < 1$, which is what we used in the last inequality.

Now, we can conclude that:

$$\sum_{k=1}^{\lfloor n^{\frac{3}{4}} \rfloor} \binom{n}{k} (1-p)^{k(n-k)} \leq \sum_{i=1}^{\lfloor \frac{n}{2} \rfloor} e^{-k(\omega(n)-2)} \leq \frac{e^{-\omega(n)+2}}{1-e^{-\omega(n)+2}} \leq 2e^{-\omega(n)+2} \to 0$$

where the last inequality holds since we assumed $\omega(n) > 10$.

– If $n^{\frac{3}{4}} \leq k \leq \lfloor \frac{n}{2} \rfloor$, we have

$$\binom{n}{k} (1-p)^{k(n-k)} \leq \left(\frac{ne}{k}\right)^k e^{-pk(n-k)}$$

Exactly as in the previous case.
Now we note that $n \geq 2k$ so we have $n-k \geq \frac{n}{2}$ and $e^{-pk(n-k)} \leq e^{\frac{-pkn}{2}}$, using this:

$$\left(\frac{ne}{k}\right)^k e^{\frac{-pkn}{2}} \leq \left(n^{\frac{1}{4}} e\right)^k e^{\frac{-k \log n}{2}} = \left(en^{-\frac{1}{4}}\right)^k$$

Where we used $k \geq n^{\frac{3}{4}}$ and $pn = \log n + \omega(n) \geq \log n$.

Now, we have

$$\sum_{k=\lceil n^{\frac{3}{4}} \rceil}^{\lfloor \frac{n}{2} \rfloor} \left(en^{-\frac{1}{4}}\right)^k \leq \frac{\left(en^{-\frac{1}{4}}\right)^k}{1-en^{-\frac{1}{4}}} \leq 2\left(en^{-\frac{1}{4}}\right)^k = 2e^{\left(1-\frac{\log n}{4}\right)k} \leq 2e^{\left(1-\frac{\log n}{4}\right)} \to 0$$

Where we assumed n is suitably large.

So we have proved:

$$\mathbb{P}(G(n,p) \text{ is disconnected}) \leq 2e^{-\omega(n)+2} + 2e^{(1-\frac{\log n}{4})} \to 0$$

So $\mathbb{P}(G(n,p)$ is disconnected$) \to 0$, as claimed.

$\square$

2.1 Asymptotic Equivalence of Erdős-Rényi and Uniform Model

The following lemma shows that uniform model and the ER-model are often interchangeable given that some not too restrictive conditions hold.

Lemma 2.2. Let $0 < p(n) < 1$ and $0 \leq a \leq 1$. If Q is a property of graphs such that for $M = Np + O(\sqrt{Np(1-p)})$ it holds that $\mathbb{P}(G(n, M) \in Q) \to a$ then $\mathbb{P}(G(n, p) \in Q) \to a$.

Proof. Let C be a large constant and define for each n

$$M'(C) = \{M : |M - Np| \leq C\sqrt{Np(1-p)}\}$$

We shall denote a random sample of $G(n, M)$ and $G(n, p)$ as G_M and G_p respectively.

Now let us define M^- to be the element of $M'(C)$ that minimizes $\mathbb{P}(G_M \in Q)$.

From the law of total probability we have:

$$\mathbb{P}(G_p \in Q) = \sum_{M=0}^{N} \mathbb{P}(G_p \in Q \mid |E(G_p)| = M)\mathbb{P}(|E(G_p)| = M) =$$

$$= \sum_{M=0}^{N} \mathbb{P}(G_M \in Q)\mathbb{P}(|E(G_p)| = M) \geq \sum_{M \in M'(C)} \mathbb{P}(G_M^- \in Q)\mathbb{P}(|E(G_p)| = M) =$$

$$\mathbb{P}(G_M^- \in Q)\mathbb{P}(G_p \in M'(C))$$

By assumption $\mathbb{P}(G_M^* \in Q) \to a$ while by using Chebyshev's inequality we have

$$\mathbb{P}(G_p \notin M'(C)) \leq \frac{\mathrm{Var}(E(G_p))}{(C\sqrt{Np(1-p)}} = \frac{\mathrm{Var}(E(G_p))}{(C\sqrt{Np(1-p)}} = \frac{1}{C^2}$$

As $E(G_p) \sim \mathrm{Bin}(N, p)$ so has variance $Np(1-p)$.

Using this we get:

$$\mathbb{P}(G_p \in Q) \geq \mathbb{P}(G_M^- \in Q)\left(1 - \frac{1}{C^2}\right)$$

In particular $\lim_{n\to\infty} \inf \mathbb{P}(G_p \in Q) \geq a(1 - \frac{1}{C^2})$.

If M^+ maximizes $\mathbb{P}(G_M \in Q)$ among $M \in M'(C)$

$$\mathbb{P}(G_p \in Q) \leq \mathbb{P}(G_{M^+} \in Q) + \mathbb{P}(G_p \notin M'(C)) \leq \mathbb{P}(G_{M^+} \in Q) + \frac{1}{C^2}$$

Hence $\lim_{n\to\infty} \sup \mathbb{P}(G_p \in Q) \leq a + \frac{1}{C^2}$.

Letting $C \to \infty$ we conclude $\mathbb{P}(G_p \in Q) \to a$ as claimed.

$\square$

We do need to impose some further restriction to get a result in the opposite direction.

This is the only direction we will explicitly use in this essay. We do use the other direction to give us better intuitive understanding of some claims. With this in mind we state a series of results of this type in both directions as the follwing lemma.

Lemma 2.2.1 Let Q be an increasing property of graphs. Assume that $M = M(n) \to \infty$ as $n \to \infty$ and that for a fixed $\delta > 0$ we have $0 \leq (1 \pm \delta)\frac{M}{N} \leq 1$. Then the following holds:

- If $\mathbb{P}(G_{\frac{M}{N}} \in Q) \to 1$, then $\mathbb{P}(G_M \in Q) \to 1$
- If $\mathbb{P}(G_{\frac{M}{N}} \in Q) \to 0$, then $\mathbb{P}(G_M \in Q) \to 0$
- If $\mathbb{P}(G_M \in Q) \to 1$, then $\mathbb{P}(G_{\frac{(1+\delta)M}{N}} \in Q) \to 1$
- If $\mathbb{P}(G_M \in Q) \to 0$, then $\mathbb{P}(G_{\frac{(1-\delta)M}{N}} \in Q) \to 0$

Proof. These statement are taken from [3] where they are proved, as we shall not use them explicitly we omit their proofs most of which go along the lines of the one we gave for **Lemma 2.2**.

□

2.2 Random Graph Processes

Let us consider the following process: we start with an empty graph on n vertices and in each step we add one of the remaining edges with equal probabilities. We stop when we have a complete graph. For example, when we are at step $M \leq N$, for every graph on M edges we have a chance of $\frac{1}{\binom{N}{M}}$ of having it and any of the remaining $N - M$ edges has a chance of $\frac{1}{N-M}$ of being chosen in the next step. We call such a process a random evolution of graph and denote it by G_0^N.

A formal mathematical definition would be to define a random evolution as a probability space on the set of $S(N)$ with each permutation being equally likely. In this case, each permutation corresponds to the order at which we have picked the edges.

We define a hitting time τ_Q for an increasing property Q as the least M, such that at step M the graph has Q.

We note that using **Lemma 2.2** above, we can relate hitting time results to the ones about threshold functions.

Corollary 2.3. p_l is a lower threshold function for Q if and only if $\lfloor p_l N \rfloor \leq \tau_Q$ almost surely while p_u is an upper threshold function for Q if and only if $\tau_Q \leq \lfloor p_u N \rfloor$ almost surely.

We leave the following theorem unproven as its proof is similar to the one given for **Theorem** 2 above, while it is very similar in nature to a result we will prove in the following section.

Theorem 2.4. Let Q be the property of being connected and R be the property of having no isolated vertices. Then $\mathbb{P}(\tau_Q = \tau_R) \to 1$.

It is clear that $\tau_Q \geq \tau_R$ as in order for a graph to be connected, it must have no isolated vertices. The remarkable thing this theorem tells us is that there is almost always an equality. Intuitively speaking, with high probability at the same time when we get rid of the last isolated vertex, the graph becomes connected.

Corollary 2.3 leads to a conclusion that connectedness and having minimal degree ≥ 1 have same threshold functions. We also observe that this fact is weaker than **Theorem 2.4** in the sense that it is implied by it, but does not imply it. It is fairly easy to prove it on its own and we note that the result we proved for the first part of **Theorem 2** is precisely this claim for lower threshold functions.

This suggests a possible strategy when dealing with harder problems, if we prove $\mathbb{P}(\tau_Q = \tau_R) \to 1$ for some properties Q, R, we can find threshold functions for one of the properties and deduce it is the same for the other. We shall use exactly this strategy in the following section.

3 Hamiltonicity

This section deals with finding the best threshold functions for the property of being Hamiltonian. The proof of the following theorem is quite hard, so we shall try to give as much intuition behind the rather technical proofs as possible.

Theorem 3. Let $\omega(n) \to \infty$. Then $p_l = \frac{\log n + \log\log n - \omega(n)}{n}$ and $p_u = \frac{\log n + \log\log n + \omega(n)}{n}$ are the lower and upper threshold functions for the property Q of being Hamiltonian.

The proof will go in two steps which differ significantly in the nature:

- We will start by showing that the above functions are threshold functions for the property R of not having a vertex of degree less than 2.
- We will show that for the respective hitting times $\mathbb{P}(\tau_Q = \tau_R) \to 1$.

From this, we will conclude using **Corollary 2.3** from last section that the result follows.

Lemma 3.1. p_l, p_u given in the statement of **Theorem** 3 are lower and upper threshold functions for the property R.

We may note that if we show p_l is a threshold function for R we can immediately conclude that p_l is a threshold function for Hamiltonicity and showing p_l is a threshold function turns out to be the harder part of this lemma.

Showing p_u is an upper threshold function for R turns out to be rather easy, but conclusion that this implies p_u is a threshold function for Hamiltonicity turns out to be the hardest of all and is given as a separate lemma.

Proof.

Let us denote the minimal degree in $G(n,p)$ by δ and the degree of node i by d_i.

Before we begin, we may notice that once again by **Lemma 2.1** we can assume $\omega(n) < \log\log\log n$. This implies that for both $p = p_l, p_u$ the **Theorem** 2 applies and implies $\mathbb{P}(G(n,p)$ is connected$) \to 1$ and consequently $\mathbb{P}(\delta \geq 1) \to 1$ and $\mathbb{P}(\delta = 0) \to 0$

$$\mathbb{P}(\delta \leq 1) = \mathbb{P}\left(\bigcup_{i=1}^{n} d_i = 1\right) + \mathbb{P}(\delta = 0, d_i \neq 1)$$

So using $\mathbb{P}(\delta = 0, d_i \neq 1) \leq \mathbb{P}(\delta = 0) \to 0$, given by **Theorem** 2, the claim is equivalent to $\mathbb{P}(\bigcup_{i=1}^{n} d_i = 1) \to 0, 1$ for $p = p_u, p_l$, respectively.

- We now prove that $p = p_u$ is an upper threshold function.

$$\mathbb{P}\left(\bigcup_{i=1}^{n} d_i = 1\right) \leq n\mathbb{P}(d_1 = 1) = n(n-1)p(1-p)^{n-2} \leq n(n-1)p(1-p)^{-2}e^{-pn} =$$

$$= (1-p)^{-2}(n-1)(\log n + \log\log n + \omega(n))e^{-\log n - \log\log n - \omega(n))} =$$

$$= (1-p)^{-2}\frac{n-1}{n}\frac{\log n + \log\log n + \omega(n)}{\log n}e^{-\omega(n)} \leq 8e^{-\omega(n)} \to 0$$

Where we assumed n is large enough to imply $2 > (1-p)^{-1}$ and $2 > \frac{\log n + \log\log n + \omega(n)}{\log n}$ which we may, as both right hand sides tend to 1.

This gives $\mathbb{P}(\delta \leq 1) \to 0$, as explained above.

- We prove that p_l is a lower threshold function.

The first idea we might try here is to define X_i as the number of vertices with degree i. Then by **Theorem** 2 $\mathbb{P}(X_0 = 0) \to 1$ and we might try to use $\mathbb{P}(X_1 = 0) \leq \frac{\mathrm{Var}(X_1)}{\mu^2}$. It turns out $\frac{\mathrm{Var}(X_1)}{\mu^2} \to 1$, so this method does not work. This method is actually equivalent to the one we present below if we used only $t = 1$ in the Bonferroni inequalities.

Let $p = \frac{\log n + \log\log n - c}{n}$ for a positive constant c.

We have by the inclusion-exclusion formula:

$$X = \mathbb{P}\left(\bigcup_{i=1}^{n} d_i = 1\right) = \sum_{k=1}^{n}(-1)^{k+1}\binom{n}{k}\mathbb{P}(d_1 = 1 \cap \cdots \cap d_k = 1)$$

Define $a_k = a_k(n) = \binom{n}{k}\mathbb{P}\left(\bigcap_{i=1}^{k} d_i = 1\right)$

We now use the Bonferroni inequalities to allow us to restrict attention only to the terms having small k. The following holds for any t:

$$\sum_{k=1}^{2t}(-1)^{k+1}a_k \leq X \leq \sum_{k=1}^{2t+1}(-1)^{k+1}a_k$$

.

Let us fix k. We now estimate a_k. We assume that n is large enough, so that at least $k < \log n$ holds.

$$a_k = \binom{n}{k} \sum_{l=0}^{\lfloor \frac{k}{2} \rfloor} \binom{k}{2l} \frac{(2l)!}{2^l} p^l (1-p)^{\binom{k}{2}-l} (1-p)^{2l(n-k)} ((n-k)p(1-p)^{n-k-1})^{k-2l}$$

This holds as for $i \leq k$ we have that within first k vertices all the edges have disjoint endpoints as $d_i = 1$.

- The sum goes over how many edges there are in the first k vertices,
- There are $\binom{k}{2l}$ subgraphs of $2l$ vertices,
- There are $\frac{(2l)!}{2^l}$ pairings among them,
- Considering there are exactly l edges among $\binom{k}{2}$, we get $p^l(1-p)^{\binom{k}{2}-l}$
- There must not be any edges joining these $2l$ vertices with the last $n-k$ vertices, which gives the factor $(1-p)^{2l(n-k)}$
- For each of the $k-2l$ non paired vertices there must be exactly one edge towards the last $n-k$ vertices, probability of this being $(n-k)p(1-p)^{n-k-1}$ for each of $k-2l$ such vertices.

After rearranging, we get:

$$a_k = \frac{n\cdots(n-k+1)}{k!}(n-k)^k p^k (1-p)^{\binom{k}{2}+k(n-k-1)} \sum_{l=0}^{\lfloor \frac{k}{2} \rfloor} \frac{k\cdots(k-2l+1)}{2^l} p^{-l}(1-p)^l (n-k)^{-2l}$$

We now want to show:

$$\sum_{l=0}^{\lfloor \frac{k}{2} \rfloor} \frac{k\cdots(k-2l+1)}{2^l} p^{-l}(1-p)^l (n-k)^{-2l} \to 1 \text{ as } n \to \infty$$

.

For this we notice $\frac{1-p}{p} \leq \frac{2n}{\log n}$ and $(n-k)^{-2l} \leq 2n^{-2l}$ hold for n large enough so

$$\sum_{l=1}^{\lfloor \frac{k}{2} \rfloor} \frac{k\cdots(k-2l+1)}{2^l} p^{-l}(1-p)^l (n-k)^{-2l} \leq \sum_{l=1}^{\lfloor \frac{k}{2} \rfloor} 2\left(\frac{k^2}{n\log n}\right)^l \leq \sum_{l=1}^{\infty} 2\left(\frac{\log n}{n}\right)^l =$$

$$2\left(\frac{1}{1-\frac{\log n}{n}} - 1\right) \to 0$$

Similarly, by bounding terms independent of l we get:

$$a_k \sim \frac{n^{2k}}{k!} p^k (1-p)^{kn} \sim \frac{n^{2k}}{k!} \frac{(\log n)^k}{n^k} e^{-pnk} = \frac{e^{kc}}{k!}$$

Where $A(n) \sim B(n)$ denotes $\frac{A(n)}{B(n)} \to 1$ as $n \to \infty$.

As the $\frac{e^{kc}}{k!}$ is independent of n, for arbitrary $\epsilon > 0$ we can get $|a_k - \frac{e^{kc}}{k!}| < \frac{\epsilon}{2t+1}$ to hold for each $k \leq 2t+1$ and n sufficiently large.

Hence, taking large enough n we get:

$$\sum_{k=1}^{2t+1}(-1)^{k+1}\frac{e^{kc}}{k!} + \epsilon \geq \sum_{k=1}^{2t+1}(-1)^{k+1}a_k \geq X \geq \sum_{k=1}^{2t}(-1)^{k+1}a_k \geq \sum_{k=1}^{2t}(-1)^{k+1}\frac{e^{kc}}{k!} - \epsilon$$

And we note further:

$$\sum_{k=1}^{2t}(-1)^{k+1}\frac{e^{kc}}{k!} \to 1 - e^{-e^c} \text{ as } t \to \infty$$

.

By choosing t large enough (and consequently n), we have:

$$\left|\mathbb{P}\left(\bigcup_{i=1}^{n} d_i = 1\right) - (1 - e^{-e^c})\right| < 2\epsilon$$

So we have:

$$\mathbb{P}\left(\bigcup_{i=1}^{n} d_i = 1\right) \to 1 - e^{-e^c} \text{ as } n \to \infty$$

We further note that **Theorem** 2 still applies so $\mathbb{P}_p\left(\bigcup_{i=1}^{n} d_i = 1\right) \to \mathbb{P}_p(\delta \leq 1)$, so in particular $\mathbb{P}_p(\delta \leq 1) \to 1 - e^{-e^c}$.

Now we note that for arbitrarily large c, $\omega(n)$ will at some point become larger than c, so using the fact that $\delta \leq 1$ is a decreasing property, we can conclude $\mathbb{P}_{p_l}(\delta \leq 1) \to 1$.

□

This completes the first part of our proof of **Theorem** 3 and as we mentioned before, this shows directly that p_l is a lower threshold function for Hamiltonicity.

It unfortunately does not imply as directly that p_u is an upper threshold function. For this we shall need the following lemma which is easily the hardest result of this essay. This is also the first result which actually uses some non-trivial graph theory.

Lemma 3.2 If Q is the property of being Hamiltonian and R the property of having $\delta \geq 2$, then for almost every graph process $\tau_Q = \tau_R$.

We shall split the proof of this result in 2 very different parts. We start by proving the graph theoretic results we shall need for the latter, more probabilistic part.

We start by introducing some notation in a given graph $G(V, E)$.

Given $U \subset V$ let $N_G(U)$ be the set of vertices connected to some element in U, but not in U, $N_G(U)$ is often referred to as the external neighbourhood of U.

We now define a notion of an expander:

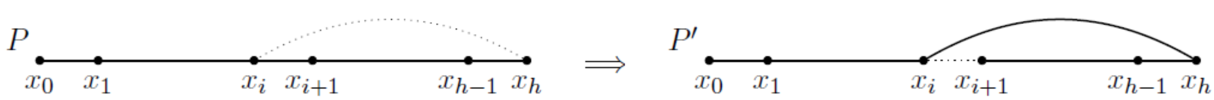
Definition. Given a positive integer k, we say a graph G is a $k-$expander if $|N_G(U)| \geq 2|U|$ for every U satisfying $|U| \leq k$.

We shall use the rotation-extension technique due to L.Posa. This is a very natural and rather simple method of finding large paths or cycles in the graph.

The basic idea is that we start with a long path and use the remaining edges to either extend it or close it to a cycle.

Let P be a simple path $x_1 x_2 \cdots x_h$, given there is an edge $x_i x_h$, an elementary rotation of P is a transformation of P to a path P' by replacing the edge $x_i x_{i+1}$ with $x_i x_h$. It might be easier to visualize this with a picture.

Picture 1. An elementary rotation.



The power of this technique lies in the following lemma:

Lemma 3.3. Given a graph $G(V, Ed)$ let P be a longest path in G. Let R be the set of all paths obtainable from P by a sequence of elementary rotations. Denote by E the set of endpoints in R and by E_- and E_+ the sets of vertices immediately preceding and following the vertices of E along P. Then $N_G(E) \subset E_- \cup E_+$

Proof. Take an $x \in E$ and $y \in V \backslash (E \cup E_- \cup E_+)$ and a path $Q \in R$ ending at x.
The lemma is equivalent to showing $xy \notin Ed$.

Let W be the set of nodes in P and consequently Q, as elementary rotations only permute the nodes of P.

If $y \in V \backslash W$, then if $xy \in Ed$ the path Q can be extended by adding y and making it longer than P, which is a contradiction.

Hence, the remaining case is if y is an element of P so $y \in W \backslash (E \cup E_- \cup E_+)$.

Let z, w denote the neighbours of y in P, where $y \notin E$ implies there are exactly 2.

Then in any path in R, z, w are still neighbours of y as an elementary rotation that removed z or w would put either y or one of z, w in E which would place y in $E \cup E_- \cup E_+$.

Now if $xy \in Ed$, an elementary rotation applied to Q produces a path in R whose endpoint is either z or w, a contradiction.

So we conclude $xy \notin Ed$ in both cases and the lemma follows.

□

The following corollary is what relates these results to our problem at hand.

Corollary 3.4. Let $G(V, Ed)$ be a graph with the longest path of length l. Assume that G does not contain a cycle of length $l + 1$ and that G is a $k-$expander. Then there are at least $\frac{k+1}{2}$ non-edges in G such that any of them would complete an $(l + 1)-$cycle in G.

Proof. Let $P = x_0x_1 \cdots x_l$ be a longest path and E, E_+, E_- be as in previous lemma.

We have $|E_-| \leq |E|$ and $|E_+| = |E| - 1$ as x_l does not have a successor in P.

So by **Lemma 3.3**, we have $|N_G(E)| \leq |E_+ \cup E_-| \leq 2|E| - 1$.

Assuming $|E| \leq k$, then as G is a $k-$expander, we have $2|E| - 1 \geq |N_G(E)| \geq 2|E|$ which is a contradiction. Therefore, $|E| > k$.

Furthermore, x_0 is an endpoint of all paths in R, so as there is no $l + 1$-cycle $x_0y \notin Ed$ for any $y \in E$ then there are $|E|$ no-edges from x_0 which complete an $(l + 1)-$cycle.

Now take $y \in E$, y is an endpoint of a path of length l in G. If we fix y instead of x_0, we can repeat the above process to conclude that there are $|E|$ no-edges from y which complete some $(l + 1)$-cycle.

By taking all $y \in E$, we conclude there are at least $\frac{|E|^2}{2} \geq \frac{(k+1)^2}{2}$ such no-edges as each no-edge was counted at most twice.

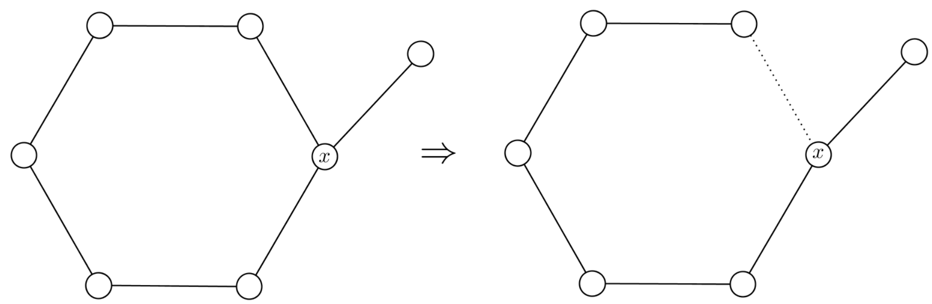
□

The reason we were considering $(l+1)-$cycles is that, provided that G is connected, adding any non-edge of the above type gives rise to an $(l + 1)$-cycle, which leads to two possibilities:

- If $l+1<n$

  Due to the connectedness of G, we must have a vertex x with an edge not in the cycle and hence, we can, by removing an x-edge in the cycle and adding the appending x-edge, increase the length of the maximal path. Picture 2 depicts this when $l=5$.

  Picture 2. How to extend an $(l+1)-$cycle to an $(l+1)-$path in the case $l=5$.



- If $l+1=n$

  The graph just became Hamiltonian. Which points to why we have considered these graph theoretic results in the first place.

We now define the edges of the above type to be called boosters.

Definition. We define a non-edge in a graph to be a booster if it either increases the length of the longest path, or makes the graph Hamiltonian.

To put our new notation to some use, we state the result we have just discussed as the following corollary.

Corollary 3.5. A connected, non-Hamiltonian, $k-$expander has at least $\frac{(k+1)^2}{2}$ boosters.

We may note that adding sequentially n boosters will bring any graph on n vertices to Hamiltonicity and this is what we were after the whole time.

This completes the rather nice graph theoretic section.

The remaining part of the proof will be quite technical, so we shall outline the idea of the proof before we start.

We consider the random process and stop it at the first time when $\delta=2$. At this time all vertices have degree ≥ 2 so it is reasonable to assume that G is with high probability a $k-$expander for some suitable k. This in fact turns out to be true.

The issue here is that while being an expander is a very good position to augment the graph to a Hamiltonian one, we can not allow ourselves to add extra edges to $G_{\tau_{\delta=2}}$, as we want to prove the graph becomes Hamiltonian exactly at the point when $\delta=2$.

We deal with this by showing that with high probability, there is a subgraph Γ_0 of G which is also a good expander and that there are almost certainly at least n boosters with respect to Γ_0 in $E(G)$. We then use **Corollary 3.5** to conclude G is Hamiltonian.

Let us define $m_1 = \frac{n \log n}{2}$ and $m_2 = n \log n$. Then by **Lemma** 3.1 and **Lemma** 2.2 almost certainly $m_1 \leq \tau_R \leq m_2$ where R is the property of having $\delta = 2$.

We define $d_0 = \lfloor \delta_0 \log n \rfloor$, where $\delta_0 > 0$ is a sufficiently small constant to be specified later.

We shall think about vertices having degree smaller than d_0 to have a "small" degree.

In the light of this, we define for a graph $G(V, E)$ on n vertices $SM(G) = \{v \in V \mid d(v) < d_0\}$.

We note that given G has $m \geq m_1$ edges, any vertex has expected degree asymptotically equal to at least $\log n$. This tells us that it is not very common for a vertex to be in $SM(G)$ and that those that are very likely to be few and far apart. We expect G to have a nice edge distribution with no small and dense vertex sets. We formalize this notions in the following lemma:

Lemma 3.6 Let $\tilde{G} = (G_i)_{i=0}^{N}$ be a random graph process on n vertices. Denote $G = G_{\tau_Q}$. Then almost certainly G has the following properties:

- **P1** Maximal degree $\Delta(G) \leq 10 \log n$.
- **P2** $|SM(G)| \leq n^{0.3}$.
- **P3** There are no vertices $x, y \in SM(G)$ such that there is a path of length at most 4 joining them, we allow $x = y$ in which case we consider paths from x to y as cycles.
- **P4** Every vertex subset $U \subset V$ of size $|U| \leq \frac{n}{\sqrt{\log n}}$ spans at most $|U|(\log n)^{\frac{3}{4}}$ edges.
- **P5** Let U be as in **P4** and let W be subset of $V \backslash U$ such that $|W| \leq |U|(\log n)^{\frac{1}{4}}$. Then the number of edges crossing between U and W is at most $\frac{d_0|U|}{2}$.
- **P6** For every pair of vertex disjoint subsets $U, W \subset V$ of size $|U| = |W| = \lceil \frac{n}{\sqrt{(\log n)}} \rceil$, G has at least $0.5n$ edges between U and W.

Proof. This is the first time we will be working in the uniform model $G(n, m)$, as this is more natural for the problem at hand.

We now give a few useful bounds on binomial coefficients mostly based on the Stirling's formula. In what follows $1 \leq x \leq k \leq n$.

I1

$$\left(\frac{n}{k}\right)^k \leq \binom{n}{k} \leq \left(\frac{en}{k}\right)^k$$

Proof. $\binom{n}{k} = \frac{n(n-1)\cdots(n-k+1)}{k!} \geq \left(\frac{n}{k}\right)^k$ follows as $\frac{n-k+i}{i} \geq \frac{n}{k}$ for each i.

While $\binom{n}{k} = \frac{n(n-1)\cdots(n-k+1)}{k!} \leq \frac{n^k}{k!} \leq \left(\frac{en}{k}\right)^k$ as Stirling's formula gives us $k! \geq \left(\frac{k}{e}\right)^k$.

I2

$$\frac{\binom{n-x}{k-x}}{\binom{n}{k}} \leq \left(\frac{k}{n}\right)^x$$

Proof. $\frac{\binom{n-x}{k-x}}{\binom{n}{k}} = \frac{(n-x)!k!}{n!(k-x)!} = \frac{k(k-1)\cdots(k-x+1)}{n(n-1)\cdots(n-x+1)} \leq \left(\frac{k}{n}\right)^x$ as $\frac{k-i}{n-i} \leq \frac{k}{n}$.

I3

$$\frac{\binom{n-x}{k}}{\binom{n}{k}} \leq e^{-\frac{kx}{n}}$$

Proof. $\frac{\binom{n-x}{k}}{\binom{n}{k}} = \frac{(n-x)!(n-k)!}{(n-k-x)!n!} = \frac{(n-k)\cdots(n-k-x+1)}{n\cdots(n-x+1)} \leq \left(\frac{n-k}{n}\right)^x \leq e^{-\frac{k}{n}x}$ where we used $\frac{n-k-i}{n-i} \leq \frac{n-k}{n}$ and $1-t \leq e^{-t}$.

We now continue to prove the statements **P1-P6** of **Lemma** 3.6.

P1 We have almost certainly $\tau_R \leq m_2$, so as **P1** is a decreasing property, it is enough to show that almost every $G = G(n, m_2)$ has **P1**.

Now in G, probability that a fixed vertex has degree $\geq 10\log n$ is at most:

$$\binom{n-1}{10\log n}\frac{\binom{N-10\log n}{m_2-10\log n}}{\binom{N}{m_2}} \leq \left(\frac{e(n-1)}{10\log n}\right)^{10\log n}\left(\frac{m_2}{N}\right)^{10\log n} \leq \left(\frac{e}{5}\right)^{10\log n} \leq n^{-5}$$

Where we have used **I1** and **I2** and $e^{\frac{3}{2}} \leq 5$.

Now, using the union bound we get $\mathbb{P}(\Delta > 10\log n) \leq n\mathbb{P}(d_i > 10\log n) \leq n^{-4} \to 0$, as desired.

P2 We note that **P2** is an increasing property, as adding edges only decreases $|SM(G)|$. So as almost surely $\tau_R \geq m_1$, it is enough to show almost every $G(n, m_1)$ has **P2**.

If $|SM(n)| \geq n^{0.3}$, then there is a subgraph of $|V_0| = k = \lceil n^{0.3} \rceil$ vertices such that number of edges crossing from V_0 to $V \backslash V_0$ is less than kd_0, as each element of V_0 has the degree at most d_0 and there are k of them in V_0.

Therefore, invoking the indicator functions on subgraphs of k vertices whether they satisfy this property, we conclude that the probability is bounded by:

$$X = \binom{n}{k} \sum_{i=0}^{d_0 k} \binom{k(n-k)}{i} \frac{\binom{N-k(n-k)}{m_1 - i}}{\binom{N}{m_1}}$$

As there are $\binom{n}{k}$ subsets of k vertices and the probability that each has the property is, after conditioning on the number of cross edges, exactly the summand (choosing i edges among $k(n-k)$ and $m_1 - i$ among the remaining $N - k(n-k)$ edges and having in total $\binom{N}{m_1}$ graphs).

$$X = \binom{n}{k} \sum_{i=0}^{d_0 k} \binom{k(n-k)}{i} \frac{\binom{N-k(n-k)}{m_1-i}}{\binom{N-i}{m_1-i}} \frac{\binom{N-i}{m_1-i}}{\binom{N}{m_1}} \leq \binom{n}{k} \sum_{i=0}^{d_0 k} \binom{kn}{i} e^{-\frac{(m_1-i)(k(n-k)-i)}{N-i}} \left(\frac{m_1}{N}\right)^i$$

Where we used $\binom{k(n-k)}{i} \leq \binom{kn}{i}$, **I3** with $n = N - i, k = m_1 - i, x = k(n-k) - i$ and **I2** with $n = N, k = m_1, x = i$. We now use **I1** twice to get:

$$X \leq \left(\frac{en}{k}\right)^k \sum_{i=0}^{d_0 k} \left(\frac{ekn}{i}\right)^i e^{-\frac{(m_1-i)(k(n-k)-i)}{N-i}} \left(\frac{m_1}{N}\right)^i \leq \left(\frac{en}{k}\right)^k \sum_{i=0}^{d_0 k} \left(\frac{eknm_1}{iN}\right)^i e^{-\frac{0.9 m_1 kn}{N}}$$

Where in the last inequality we used:

$$e^{-\frac{(m_1-i)(k(n-k)-i)}{N-i}} \leq e^{-\frac{0.9 m_1 kn}{N}} \Leftrightarrow$$

$$\frac{0.9 m_1 kn}{N} \leq \frac{(m_1 - i)(k(n-k) - i)}{N - i}$$

Which is implied by $0.9 m_1 kn \leq (m_1 - i)(k(n-k) - i) \Leftrightarrow k^2 m_1 + i(k(n-k) + m_1 - i) \leq 0.1 m_1 kn$ which holds as $LHS = O(n^{1.6} \log n)$ while $RHS = cn^2 \log n$ so this will clearly hold for n large enough.

We now claim $\left(\frac{eknm_1}{iN}\right)^i \leq \left(\frac{eknm_1}{d_0kN}\right)^{d_0k}$, which follows as by differentiating $\left(\frac{c}{x}\right)^x$ we get this is increasing in x provided $c \geq ex$ and this holds in particular as in our case $\frac{eknm_1}{N} \geq ei$ holds for $\delta_0 < 0.5$. Using this in our expression and summing up, we get:

$$X \leq \left(\frac{en}{k}\right)^k (d_0k+1) \left(\frac{eknm_1}{d_0kN}\right)^{d_0k} e^{-\frac{0.9m_1kn}{N}} = (d_0k+1) \left(\frac{en}{k}\left(\frac{en\log n}{d_0(n-1)}\right)^{d_0} e^{-\frac{0.9m_1n}{N}}\right)^k \leq$$

$$\leq 2d_0k\left(\frac{3n}{k}\left(\frac{3}{\delta_0}\right)^{\delta_0\log n} e^{0.8\log n}\right)^k \leq 2d_0k\left(3n^{0.7}e^{0.05\log n}e^{0.8\log n}\right)^k \leq$$

$$\leq 4\delta_0 \log n n^{0.3}(3n^{-0.05})^k \to 0$$

Where only non trivial inequality we used here was that we can pick δ_0 small enough such that $\left(\frac{3}{\delta_0}\right)^{\delta_0} \leq e^{0.05}$, which follows as LHS tends to 1 as $\delta_0 \to 0$.

This completes the proof for this property.

P3 Since with high probability $m_1 \leq \tau_R \leq m_2$, it is enough to prove the following: For almost every $G(n, m_2)$, every two (possibly identical) vertices of $SM(G(n, m_2))$ are not connected by a path of length at most 4 in $G_(n, m_2)$.

Let us first prove that almost certainly there is no such path in $G_(n, m_1)$.

We only treat the case where the endpoints of the path are distinct, as the remaining case is very similar.

Fix $1 \leq r \leq 4$, a sequence P of distinct vertices $v_0, \cdots v_r$ in V and denote by A_P the event $v_iv_{i+1} \in E(G(n, m_1))$ for every $0 \leq i \leq r-1$. We now have using **I2**:

$$\mathbb{P}(A_P) = \frac{\binom{N-r}{m_1-r}}{\binom{N}{m_1}} \leq \left(\frac{m_1}{N}\right)^r = \left(\frac{\log n}{n-1}\right)^r$$

We now condition on A_P, so in order for both v_0, v_r to fall into $SM(G)$ out of the total of $2(n-2)$ edges between $\{v_0, v_r\}$ and the rest of the graph at most $2(d_0-1)$ are present. Hence:

$$\mathbb{P}(v_0, v_r \in SM(G) \mid A_P) \leq \sum_{i=0}^{2d_0-2} \binom{2n-4}{i} \frac{\binom{N-r-2n+4}{m_1-r-i}}{\binom{N-r}{m_1-r}} \leq (2d_0-1)\binom{2n-4}{2d_0-2} \frac{\binom{N-r-2n+4}{m_1-r-2d_0+2}}{\binom{N-r}{m_1-r}} \leq$$

$$\leq 2d_0 \binom{2n-4}{2d_0-2} \frac{\binom{N-r-2n+4}{m_1-r-2d_0+2}}{\binom{N-r-2d_0+2}{m_1-r-2d_0+2}} \frac{\binom{N-r-2d_0+2}{m_1-r-2d_0+2}}{\binom{N-r}{m_1-r}} \leq$$

$$\leq 2d_0 \left(\frac{en}{d_0-1}\right)^{2d_0-2} e^{-\frac{(m_1-r-2d_0+2)(2n-2d_0-2)}{N-r-2d_0+2}} \left(\frac{m_1-r}{N-r}\right)^{2d_0-2} \leq$$

$$\leq 2d_0 \left(\frac{em_1 n}{(d_0-1)N}\right)^{2d_0-2} e^{-\frac{1.9m_1 n}{N}} \leq n^{-1.8}$$

Where all these inequalities come from very similar facts to those in the proof of **P2**.

Now applying the union inequality over all such $r+1$ vertices we conclude:

$$\mathbb{P}(G(n, m_1) \text{ does not have P3}) \leq \sum_{r=1}^{4} n^{r+1} \left(\frac{\log n}{n-1}\right)^r n^{-1.8} \to 0$$

.

Therefore, we can assume that at time m_1 of the random graph process the current graph does not have a forbidden short path between the vertices of $SM(G)$.

If an edge $m_1 \leq i \leq m_2$ closes a short path between vertices in $SM(G(n, m_1))$, then it should fall into the set C of edges at distance ≤ 3 from $SM(G(n, m_1))$.

Furthermore, we can assume, almost certainly, by **P2** that $|SM(G(n, m_1))| \leq n^{0.3}$ and by **P1** the maximal degree of $G(n, m_2)$ is $\leq 10 \log n$.

So $|C| \leq |SM(G(n, m_1))|(10 \log n)^3 \leq n^{0.3}(10 \log n)^3$.

We also note that probability of placing edge i inside C is at most $\frac{\binom{|C|}{2}}{N-m_1}$, as we have at least $N - m_1$ edges left and there are $\binom{|C|}{2}$ edges in C.

Taking the union bound over such i $\mathbb{P}(\text{P3 holds in } G(n, m_2)) \leq (m_2 - m_1)\frac{\binom{|C|}{2}}{N-m_1} = o(n^{-0.3}) \to 0$ as desired.

P4-P6 The proofs for **P4-P6** are very similar to the ones given above, so we omit them here.

□

From now on, we can assume that G_{τ_R} has properties **P1-P6**.

We aim to prove G_{τ_R} contains a relatively small subgraph Γ_0 that is a $\frac{n}{4}-$expander and that in $E(G)\backslash E(\Gamma)$ there are at least n boosters.

We form a random subgraph of $G(V, E)$ as follows: for every $v \in V\backslash SM(G)$ choose a set $E(v) \subset E$ of size d_0 incident to v uniformly at random. For $v \in SM(G)$ define $E(v)$ to be all edges in G incident to v.

Finally, we define Γ_0 to be the subgraph of G having $E(\Gamma_0) = \bigcup_{v\in V} E(v)$.

Intuitively, we take all edges touching $SM(G)$ and sparsify the remaining edges. We now prove that Γ_0 is almost certainly a $\frac{n}{4}-$expander.

Lemma 3.7. With high probability (over the choices of $E(v)$) the subgraph Γ_0 of a graph G satisfying **P1-P6** is a $\frac{n}{4}-$expander with at most $\frac{d_0 n}{2}$ edges.

Proof. We first note that we will have $\Delta(\Gamma_0) = d_0$, which gives us that $E(\Gamma_0) \leq \frac{nd_0}{2}$, as each vertex contributes at most d_0 edges and each edge is counted twice.

P7 For every pair of disjoint sets U, W of size $|U| = |W| = \left\lceil \frac{n}{\sqrt{\log n}} \right\rceil$, Γ_0 has at least one edge between U and W.

Fix sets U, W as above. Ee know by **P6** that G has at least $0.5n$ edges between U and W.

For a vertex $u \in U$, the probability that no edge falls into $E(u)$ is at most:

$$\frac{\binom{d_G(u)-d_G(u,W)}{d_0}}{\binom{d_G}{d_0}} \leq e^{-\frac{d_G(u,W)d_0}{d_G(u)}} \leq e^{-\frac{d_G(u,W)d_0}{10\log n}}$$

Where we used **I3** and **P1** and $d_G(u)$ is degree of u in G, while $d_G(u, W)$ is the number of edges from u to W.

Now the probability of having no edge between U and W is at most:

$$\bigcap_{u\in U} e^{-\frac{d_G(u,W)d_0}{10\log n}} = e^{-\frac{d_0}{10\log n}\sum_{u\in U} d_G(u,W)} = e^{-\frac{d_0}{10\log n}E_G(U,W)} \leq e^{-\frac{d_0 n}{20\log n}}$$

Where $E_G(U, W)$ is the number of edges between U and W in G for which **P6** implies is larger than $0.5n$.

We denote $l = |U| = |W| = \frac{n}{\sqrt{\log n}}$. Now applying the union inequality over all U, W we get using **I1** and **I3**

$$\mathbb{P}(\Gamma_0 \text{ does not have P7}) \leq \binom{n}{|U|}\binom{n-|W|}{|W|}e^{-\frac{d_0 n}{20\log n}}$$

$$\leq \binom{n}{l}^2 e^{-\frac{l^2}{n}} e^{-\frac{d_0 n}{20\log n}} \leq \left(\frac{en}{l}\right)^{2l} e^{-\frac{n}{\log n}-\frac{d_0 n}{20\log n}} =$$

$$= (e^2\log n)^l e^{-\frac{(d_0+20)n}{20\log n}} = \left(\log n e^{2-\frac{d_0+20}{20\sqrt{\log n}}}\right)^{\frac{n}{\sqrt{\log n}}} \to 0$$

Where we used $\log n e^{2-\frac{d_0+20}{20\sqrt{\log n}}} \leq e^2 \log n e^{-\frac{\delta_0\sqrt{\log n}}{40}} \to 0$.

This proves that almost every Γ_0 has **P7**.

We now show that for a given graph G satisfying **P2,P3,P4,P5,P7**, any subgraph Γ_0 is an $n/4-$expander.

Let S be a subset of V of size $|S| \leq \frac{n}{4}$.

Denote $S_1 = S \cap SM(G)$ and $S_2 = S \backslash S_1$.

- Case $|S_2| \leq \frac{n}{\sqrt{\log n}}$

  We have $\delta(\Gamma_0) \geq 2$ and each pair of vertices in $SM(G)$ are at least 4 edges apart, so $N_{\Gamma_0}(S_1) \geq 2|S_1|$ as no neighbours of S_1 can be shared (this would imply there is a path of length 2).

  Vertices from S_2 have degree at least d_0 in Γ_0. Also by **P4** we have $E_{\Gamma_0}(S_2) \leq E_G(S_2) \leq |S_2|(\log n)^{3/4}$ and therefore, at most this much in Γ_0 as well. It thus follows:

$$E_{\Gamma_0}(S_2, V - S_2) > d_0|S_2| - 2E_{\Gamma_0}(S_2) > (d_0 - 2(\log n)^{3/4})|S_2| > \frac{1}{2}d_0|S_2|$$

  Hence, we can conclude by setting $U = S_2$ and $W = N_{\Gamma_0}(S_2)$ that we can conclude by **P5** that $|N_{\Gamma_0}(S_2)| \geq |S_2|(\log n)^{\frac{1}{4}}$.

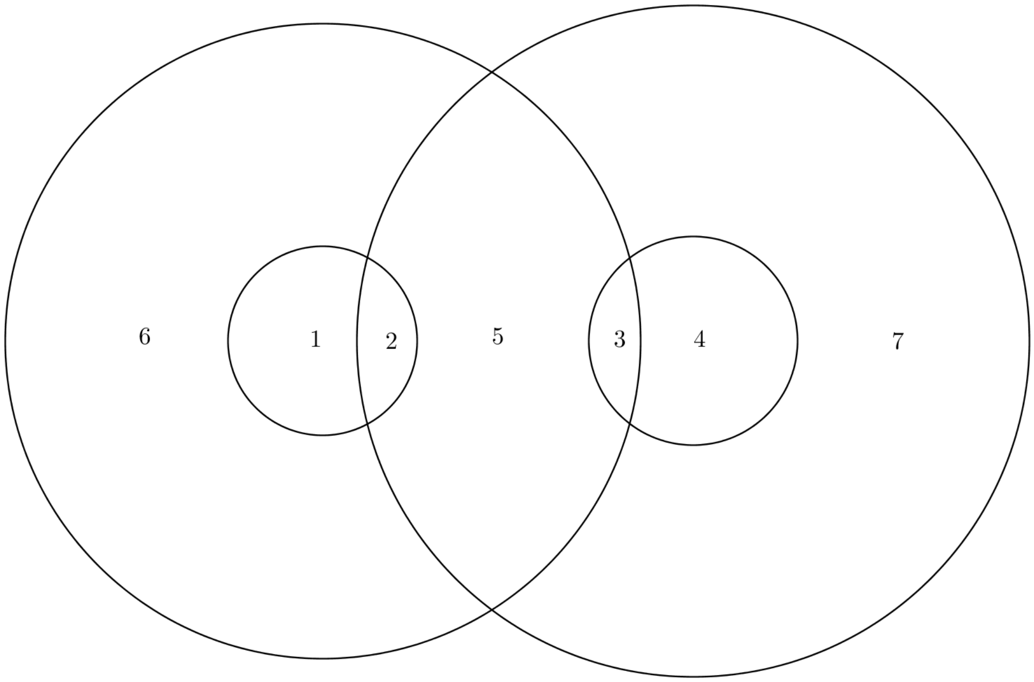
  Finally, due to **P3** we have $S_1 \cap N_{\Gamma_0}(S_2) \leq |S_2|$, as each element of S_1 can have at most 1 neighbour in S_2, as otherwise we would have a path of length 2 connecting vertices of $SM(G)$, contradicting **P3**.

  Now, we notice:

$$N_{\Gamma_0}(S) = N_{\Gamma_0}(S_2) \cup (N_{\Gamma_0}(S_1) - (N_{\Gamma_0}(S_2) \cup S_2)) - (S_1 \cap N_{\Gamma_0}(S_2))$$

  It may help to visualize this as a picture:

Picture 3. The sets S_1, S_2 with their external neighborhoods



In the picture $S_1 = 1 \cup 2$, $S_2 = 3 \cup 4$, $N_{\Gamma_0}(S_1) = 3 \cup 5 \cup 6$, $N_{\Gamma_0}(S_2) = 2 \cup 5 \cup 7$.

Also, the above set equality is equivalent to

$$5 \cup 6 \cup 7 = ((2 \cup 5 \cup 7) \cup ((3 \cup 5 \cup 6) - (2 \cup 3 \cup 4 \cup 5 \cup 7))) - ((1 \cup 2) \cap (2 \cup 5 \cup 7))$$

, which does hold.

Now $N_{\Gamma_0}(S) = |N_{\Gamma_0}(S_2)| + |(N_{\Gamma_0}(S_1) - (N_{\Gamma_0}(S_2) \cup S_2))| - |(S_1 \cap N_{\Gamma_0}(S_2))| \geq |S_2|(\log n)^{\frac{1}{4}} + 2|S_1| - |S_2| \geq 2(|S_1| + |S_2|) = 2|S|$ for n large enough to make $(\log n)^{\frac{1}{4}} - 1 \geq 2$.

- Case $\frac{n}{\sqrt{\log n}} \leq |S_2| \leq \frac{n}{4}$

  We notice that if $|V - N_{\Gamma_0}(S_2)| > \frac{n}{\sqrt{\log n}}$, we can choose U to be any subset of S_2 and W any subset of $V - N_{\Gamma_0}(S_2)$ of suitable size in $P7$ which gives a contradiction, as we must have an edge between U, W so an element of W would be in $V - N_{\Gamma_0}(S_2)$.

  Now using the same inequality as in the previous case:

$$N_{\Gamma_0}(S) \geq |N_{\Gamma_0}(S_2)| + 2|S_1| - |S_2| \geq n - \frac{n}{\sqrt{\log n}} - \frac{n}{4} \geq \frac{n}{2} \geq 2|S|$$

This completes our proof of **Lemma** 3.7.

□

Proposition 3.8. Every $\frac{n}{4}-$expander Γ on n vertices is connected.

Proof. Assume Γ is disconnected.

Then it has a component C of size $|C| \leq \frac{n}{2}$.

Pick some subset U of C with size $|U| = \min(\lfloor \frac{n}{4} \rfloor, |C|)$, so the neighborhood of U is of size at least $2|U|$ and falls entirely within C. Hence, $|U| \neq |C|$ so $|U| = \lfloor \frac{n}{4} \rfloor$, but we must have $|C| \geq 3|U|$ as both U and $N_G(U)$ are in C and are disjoint, so we have our contradiction.

□

We now proceed to find suitable boosters:

Lemma 3.9. Let $G \equiv G_{\tau_R}$. Assume the constant δ_0 is small enough. Then almost certainly, for every $\frac{n}{4}-$expander $\Gamma \subset G$ on the same vertex set with $|E(|\Gamma|)| \leq d_0 n + n$ is Hamiltonian, or G contains at least one booster with respect to Γ.

Proof. Consider some Γ that violates the condition of the lemma.

Then, Γ is $\frac{n}{4}-$expander, non-Hamiltonian and by **Lemma** 3.8 it is connected.

Therefore, we can apply **Corollary** 3.5 to conclude there are at least $\frac{\left(\frac{n}{4}\right)^2}{2} = \frac{n^2}{32}$ boosters with respect to $|\Gamma|$ and no such booster is in G.

We have almost certainly $m_1 \leq \tau_R \leq m_2$, so the probability of having a γ violating the above conditions is less than

$$o(1) + \sum_{m=m_1}^{m_2} \sum_{i \leq d_0 n + n} \frac{\binom{n}{i}\binom{N-i-\frac{n^2}{32}}{m-i}}{\binom{N}{m}}$$

Where we sum over relevant values of m, while $o(1)$ term accounts for the probability of τ_R falling outside of the interval $[m_1, m_2]$. Then, we sum over the possible values of $|E(\Gamma)|$. Then, we bound from above the number of $\frac{n}{4}-$expanders with i edges by $\binom{n}{i}$ and we require that edges of Γ are present in $G(n,m)$, while at least $\frac{n^2}{32}$ boosters have to be omitted, which accounts for the last term.

We now proceed to bound the summands above using **I1,I2,I3**.

$$\binom{n}{i} \frac{\binom{N-i-\frac{n^2}{32}}{m-i}}{\binom{N}{m}} \leq \left(\frac{en}{i}\right)^i \frac{\binom{N-i-\frac{n^2}{32}}{m-i}}{\binom{N-i}{m-i}} \frac{\binom{N-i}{m-i}}{\binom{N}{m}} \leq \left(\frac{en}{i}\right)^i e^{-\frac{\frac{n^2}{32}(m-i)}{N-i}} \left(\frac{m}{N}\right)^i$$

Pickig δ_0 small enough, we have $\frac{m}{17} \leq \frac{m-i}{16}$, we also note that $2N - 2i = n(n-1) - 2i \leq n^2$, so using these inequalities we get the above expression is at most:

$$\left(\frac{em}{i}\right)^i e^{-\frac{m}{17}} \leq \left(\frac{em}{d_0 n + n}\right)^{d_0 n + n} e^{-\frac{m}{17}} \leq \left(\left(\frac{3}{\delta_0}\right)^{2\delta_0}\right)^{n \log n} e^{-\frac{m}{17}} \leq \left(e^{\frac{1}{68}}\right)^{n \log n} e^{-\frac{n \log n}{34}} = n^{-\frac{n}{68}}$$

Where the first inequality comes from the same reasoning we gave in the proof of **P1,P2** and the second comes from $\frac{n \log n}{2} \leq m \leq n \log n$ and $\left(\frac{3}{\delta_0}\right)^{2\delta_0} \to 1$ as $\delta_0 \to 0$.

And the initial sum is smaller than $(m_2 - m_1 + 1)(d_0 n + d_0)n^{-\frac{n}{68}} \leq 2n \log n n^{-\frac{n}{68}} \to 0$ as desired.

This completes the proof of the lemma.

□

It also completes the proof of **Theorem** 3, but considering the legth of the proof we give a summary.

We have shown that almost every graph $G = G_{\tau_R}$ satisfies conditions of **Lemma** 3.7, so it has an $\frac{n}{4}-$expander Γ_0 as a subgraph of G.

Now we keep adding boosters which are in $E(G)$ to this expander, until the current graph becomes Hamiltonian.

The event that we never reach Hamiltonicity, corresponds to there being an expander with $\leq d_0 n + n$ edges and no boosters in G, so **Lemma** 3.9 implies this happens almost never, so indeed G is almost certainly Hamiltonian.

4 Giant Component

In this section we focus on the properties of random evolution of graphs. We are mostly interested in how the component structure changes with time.

The focus of thi section is on the fact that in a random process initially all the components have size $O(\log n)$ all up to the point when suddenly one component starts to dominate and increases from $O(\log n)$ to $O(n)$ while the remaining components stay of $O(\log n)$. This component is called the giant component and we shall prove its existence and state some results about it.

We start by describing how the component structure evolves with time.

Let us consider $G(n, M)$ with $M = \frac{cn}{2}$.

- When c is small almost surely every component of G is a tree or unicyclic and it has size $O_c(\log n)$.
- As c grows towards 1 the size of components is growing but largest components merge mostly with small trees and therefore, grow slowly and quite smoothly. During this period all the components are still of $O_c(\log n)$.
- At some point large components become large enough so that instead of accumulating small components they start merging together. This happens quite suddenly around time corresponding to $c = 1$. As two large components merge, their size increases significantly which increases the probability of that component being further merged soon after. Fairly quickly all large components merge into a single giant component. We call this period a phase transition.
- as $c > 1$ the giant component dominates and is of size $O_c(n)$ while all the remaining components have size $O(\log n)$.

The remarkable thing about this is that $G(n, \lfloor 0.499n \rfloor)$ and $G(n, \lceil 0.501n \rceil)$ have significantly different component structure.

We also note that from **Theorem** 2, by applying **Lemma** 2.2.1 to switch between models, at time $M = \frac{n \log n}{2}$ the last isolated vertex joins the giant and G becomes connected.

We now state the main theorem of this section, we shall work in the ER-model as independence of edges makes it simpler to work with. We can still, by defining suitable properties, switch to the uniform model by invoking **Lemma** 2.2.1.

Theorem 4. Let $p = \frac{c}{n}$, where $c > 0$ is a constant.

- If $c < 1$ then almost surely the largest component of $G(n, p)$ has at most $\frac{8}{(1-c)^2} \log n$ vertices.
- If $c > 1$, $G(n, p)$ contains, almost surely, a component of size $(1 + o_p(1))\alpha n$, where α is a unique solution to $e^{-\alpha c} = 1 - c$ in $(0, 1)$. Also the size of the second largest component is almost certainly at most $\frac{16}{(1-c)^2} \log n$

Where by $x_n = o_p(1)$ we mean: given any $\epsilon > 0$ we have with high probability $x_n < \epsilon$.

Proof. We shall exploit a certain similarity with branching processes so we start by stating a few standard results.

Let Z_n be the size of $n-$th generation, so Z_0=1 and Z_n is a sum of Z_{n-1} independent random variables with same distribution as Z_1 called hereafter offspring distribution. We also define $Z = \sum_{i \geq 0} Z_i$ to be the total number of offspring in all generations.

We define $f(x) = \sum_{i=0}^{\infty} x^i \mathbb{P}(Z_1 = i)$ to be the generating function of Z_1.

The probability of extinction is

$$\rho = \mathbb{P}(Z < \infty) = \lim_{n \to \infty} \mathbb{P}(Z_n = 0)$$

Lemma 4.1.1 Using the above notation:

- If $\mathbb{E}\, Z_1 \leq 1$ we have $\rho = 1$ unless $\mathbb{P}(Z_1 = 1) = 1$.
- If $\mathbb{E}\, Z_1 > 1$ and $\mathbb{P}(Z_1 = 0) > 0$ we have ρ to be the smallest solution to $f(x) = x$ in $(0, 1)$.

Proof. Omitted.

□

We now consider the relation of extinction probabilities of branching processes with Binomial and Poisson distributions.

If $Z_1 \sim \text{Poiss}(c)$ then

$$f_{Z_1}(x) = \sum_{i=0}^{\infty} \frac{x^i c^i}{i!} e^{-c} = e^{xc-c}$$

So given $\mathbb{E}\, Z_1 = c > 1$ the probability of extinction is determined by $\rho = e^{c(x-1)}$ and if we denote the probability of ultimate survival as $\alpha = 1 - \rho$ we have $\alpha + 1 = e^{-\alpha c}$

Proposition 4.1.2. The equation $h(x) = x + 1 - e^{xc} = 0$ always has a unique solution in $(0, 1)$ provided $c > 1$.

Proof. We have $h(0) = 1$ while $h(1) = 2 - e^c < 0$ as $c > 1$ so by the intermediate value theorem there is always ate least one solution.

Considering that $h' = 1 - ce^{xc} < 0$ on $[0, 1]$ function h is strictly decreasing on $(0, 1)$ so it has indeed exactly one solution on this interval.

□

If $Y_n \sim \text{Bin}(n, p)$ where $np \to c > 1$ as $n \to \infty$

$$f_{Y_n}(x) = \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} x^i = (1 - p + xp)^n$$

For a fixed x we have

$$\lim_{n\to\infty} f_{Y_n}(x) = \lim_{n\to\infty} \left(1 + \frac{c(x-1)}{n}\right)^n = e^{xc-c} = f_{Z_1}(x)$$

Therefore f_{Y_n} converge pointwise to f_{Z_1}.

While this is not enough to show the following lemma, using **Lemma 4.1.1**, **Proposition 4.1.1** and being a bit more careful with our analysis we have

Lemma 4.1.3 With the notation as above $\rho_{Y_n} \to \rho_{Z_1}$.

Proof. Omitted.

We now turn to give some intuition for our proof of the theorem:

We consider the following process, often called breadth first search.

We start with a vertex v of G at each step we have 3 sets A, B, C such that initially $A = \{\}$, $B = \{v\}$ and $C = V \backslash \{v\}$.

At each step we pick a vertex x in B which shortest path to v is smallest and add all its neighbours currently in C to B and move x to A.

If after t steps B is empty, current A is the component of v in G.

This process resembles a branching process but unlike in the branching process where the number of immediate offspring of an element does not depend on the past here the number of nodes $X_i = X_i(n, m, p)$ we add at time i provided m of its nodes have already been found (are in $A \cup B$) has distribution $X_i \sim \text{Bin}(n - m, p)$ which is dependant on the past.

Regardless of this, if m is small enough this process can be approximated by a branching process with distribution $\text{Bin}(n, p)$. So we might expect that the probability a vertex is in a small component is given by the probability of extinction which happens with probability one given $c < 1$. If on the other hand $c > 1$ then the process continues indefinitely with some positive probability $1 - \rho_c$ so we might expect there to be $(1 - \rho_c + o(1))n$ vertices in its component.

We now formalize this idea as follows:

- Case $pn = c < 1$

  Probability that a vertex v belongs to a component of size $k(n)$ is smaller than the probability that the sum of k random variables X_i, as above, is at least $k-1$.

  We can also bound X_i by independent random variable $Y_i \sim \text{Bin}(n,p)$.

  Furthermore, as X_i $S = S(k,n,p) = \sum_{i=1}^{k} Y_i \sim \text{Bin}(kn,p)$.

  We can now conclude, by using the union bound on all n vertices, that probability of having a component of size $k \geq \frac{8\log n}{(1-c)^2}$ is bounded by

$$n\mathbb{P}(S \geq k-1) = n\mathbb{P}(S \geq ck + (1-c)k - 1)$$

  In order to bound this we prove the following inequality:

  Proposition 4.2. If random variable $X \sim \text{Bin}(n,p)$ then for $t \geq 0$

$$\mathbb{P}(X \geq \mathbb{E}\,X + t) \leq e^{-\frac{t^2}{2(np+\frac{t}{3})}}$$

  Proof. Let us define $u = \log(np+t) + \log(1-p) - \log(n-np-t) - \log p$. We do note that have we left u undefined we would have found this to be the minimum in one of the following inequalities.

$$\mathbb{P}(X \geq \mathbb{E}\,X + t) = \mathbb{P}\left(e^{uX} \geq e^{u(\mathbb{E}\,X+t)}\right) \leq e^{-u(\mathbb{E}\,X+t)}\,\mathbb{E}\left(e^{uX}\right) = e^{-u(np+t)}(1-p+pe^u)^n =$$

$$= \left(\frac{np}{np+t}\right)^{np+t}\left(\frac{n-np}{n-np-t}\right)^{n-np-t} = e^{-npg\left(\frac{t}{np}\right)-n(1-p)g\left(\frac{-t}{n(1-p)}\right)}$$

  Where we used Markov's inequality for e^{uX} and $e^{u(\mathbb{E}\,X+t)}$ and defined $g(x) = (x+1)\log(x+1) - x$ for $x > -1$ and assumed $t < n(1-p)$

  We note $g(x)$ is decreasing when $x \leq 0$ and increasing on $x \geq 0$ so it minimum is at $x = 0$ $g(0) = 0$, hence $g(x) \geq 0$ and

$$\mathbb{P}(X \geq \mathbb{E}\,X + t) \leq e^{-npg\left(\frac{t}{np}\right)}$$

  For which we may also note that holds trivially given $t > n(1-p)$. We now want to prove

$$npg\left(\frac{t}{np}\right) \geq \frac{t^2}{2(np+\frac{t}{3})} \Leftrightarrow$$

$$g(x) \geq \frac{x^2}{2+\frac{2x}{3}} = h(x)$$

Where $x = \frac{t}{np}$. Now noticing $g(0) = g'(0) = 0 = h'(0) = h(0)$ and

$$g''(x) = \frac{1}{1+x} \geq \frac{1}{\left(1+\frac{x}{3}\right)^3} = h''(x)$$

so $g(x) \geq h(x)$ holds indeed and this completes the proof of the lemma.

□

Returning to the **Theorem** 4 we apply this proposition to our case to conclude

$$n\mathbb{P}(S \geq ck+(1-c)k-1) \leq ne^{-\frac{((1-c)k-1)^2}{2\left(ck+\frac{(1-c)k-1}{3}\right)}} \leq ne^{-\frac{(1-c)^2k^2}{4\left(\frac{2ck+k}{3}\right)}} \leq ne^{-\frac{(1-c)^2k}{4}} \leq ne^{-2\log n} = n^{-1} \to 0$$

Which completes this part of the proof.

- Case $pn = c > 1$

  Let us define $k_- = \frac{16}{(1-c)^2}\log n$ and $k_+ = n^{\frac{2}{3}}$

  We first show that with high probability for every vertex v and every k such that $k_- \leq k \leq k_+$ the process described above starting from v either dies out before k_-, or at the $k-$th step there are at least $\frac{(c-1)k}{2}$ elements in the set B implying there are no components with size $k_- \leq k \leq k_+$.

Lemma 4.3. With k_- and $_+$ defined as above, with high probability there is no component of size k for which $k_- \leq k \leq k_+$ holds.

Proof. Now we can bound each X_i for $1 \leq i \leq k$ from below by $X_i^- \sim \text{Bin}\left(n - \frac{c+1}{2}k_+, p\right)$ which are mutually independent. If we denote by A_i, B_i sizes of the sets A, B at step i respectively we have:

$$\mathbb{P}\left(B_k \leq \frac{(c-1)k}{2}\right) = \mathbb{P}\left(A_k + B_k \leq k + \frac{(c-1)k}{2}\right) =$$

$$= \mathbb{P}\left(\sum_{i=1}^{k} X_i \leq k + \frac{(c-1)k}{2}\right) \leq \mathbb{P}\left(\sum_{i=1}^{k} X_i^- \leq \frac{(c+1)k}{2}\right)$$

Applying the union bound on this happening to a vertex at time $k_- \leq k \leq k_+$ we conclude that probability of this happening for any component is at most:

$$P = n \sum_{k=k_-}^{k_+} \mathbb{P}\left(\sum_{i=1}^{k} X_i^- \leq \frac{(c+1)k}{2} \right)$$

We now have $\sum_{i=1}^k X_i^- \sim \text{Bin}\left(kn - k\frac{c+1}{2}k_+, p\right)$

Proposition 4.4 If $X \sim \text{Bin}(n,p)$ then for $t \geq 0$

$$\mathbb{P}(X \leq \mathbb{E}\, X - t) \leq e^{-\frac{t^2}{2np}}$$

Proof. As the proof goes along the same lines as that of **Proposition 4.2** we omit it here, the proof can be found as (2.6) in [3].

□

Applying this inequality in our case $\mathbb{E}(\sum_{i=1}^k X_i^-) = knp - \frac{c+1}{2}kpk_+ = kc - \frac{c+1}{2}kpk_+$ we set $t = kc - \frac{c+1}{2}kpk_+ - \frac{(c+1)k}{2} = \frac{(c-1)k}{2} - \frac{c+1}{2}kpk_+$ to get:

$$P \leq n \sum_{k=k_-}^{k_+} e^{-\frac{\left(\frac{(c-1)k}{2} - \frac{c+1}{2}kpk_+\right)^2}{2kc-(c+1)kpk_+}}$$

We now proceed to bound the exponent by noticing $pk_+ = cn^{-\frac{1}{3}} \to 0$ (this also implies that $t > 0$ for n large enough, what we required above).

$$\frac{\left(\frac{(c-1)k}{2} - \frac{c+1}{2}kpk_+\right)^2}{2kc - (c+1)kpk_+} = \frac{k\left(\frac{(c-1)}{2} - \frac{c+1}{2}pk_+\right)^2}{2c - (c+1)pk_+} \geq \frac{k\left(\frac{(c-1)}{\frac{3}{\sqrt{2}}}\right)^2}{2c} = \frac{k(c-1)^2}{9}$$

Using this above we get

$$P \leq n \sum_{k=k_-}^{k_+} e^{-\frac{k(c-1)^2}{9}} \leq nk_+ e^{-\frac{k_-(c-1)^2}{9}} = n^{\frac{5}{3}} e^{-\frac{16}{9}\log n} = n^{-\frac{1}{9}} \to 0$$

This completes the proof of the claim that almost certainly there are no "medium"-sized components.

□

Lemma 4.5 With high probability any 2 vertices in components of size at least k_+ are connected.

textbfProof. This lemma tells us that with high probability there is only one "large" component.

Given two vertices v, v' both in some components of size at least k_+ we denote by P the probability they are in the same component.

We run the process described above first from v and then from v'.

In proving **Lemma 4.3** we have shown there are at least $\frac{(c-1)k_+}{2}$ vertices in sets B for both v and v' therefore the chance of v and v' being in distinct components is

$$1 - P \leq (1-p)^{\left(\frac{(c-1)k_+}{2}\right)^2} = \left(1 - \frac{c}{n}\right)^{\frac{(c-1)^2 n^{\frac{4}{3}}}{4}} \leq e^{-\frac{c(c-1)^2 n^{\frac{1}{3}}}{4}}$$

Where we used $1 - p \leq e^{-p}$ and np.

Now taking the union bound over all pairs of vertices in "large" components we get that probability of having two in different components is less than

$$\frac{n(n-1)}{2} e^{-\frac{c(c-1)^2 n^{\frac{1}{3}}}{4}} \to 0$$

This completes the proof of our lemma.

$\square$

This lemma shows that with high probability there is only one "large" component. So the second largest component must be "small" so in particular size of the second largest component is at most $k_- = \frac{16}{(1-c)^2} \log n$ as claimed in the theorem.

We are therefore left with estimating the size of the largest component. We shall call a vertex small if it is in a component of size $\leq k_-$.

Lemma 4.6. For any $\epsilon > 0$ there are with high probability less than $(1+\epsilon)(1 - \alpha + o_p(1))n$ small vertices in $G(n,p)$.

Proof. We note that probability $\rho(n,p)$ that a vertex is small is bounded above by extinction probability $\rho_+(n,p)$ for a branching process with offspring distribution $\mathrm{Bin}(n-k_-, p)$ and bounded below by extinction probability $\rho_-(n,p) + o(1)$ for a branching process with offspring distribution $\mathrm{Bin}(n,p)$ where $o(1)$ term is here to account for the case of the process lasting more than k_- steps.

Therefore, we conclude as both $np = c \to c$ and $(n - k_-)p = c\left(1 - \frac{16 \log n}{n(1-c)^2}\right) \to c$ that $\rho(n,p) = 1 - \alpha + o(1)$ where alpha is the survival probability of a branching process with offspring distribution $\mathrm{Poiss}(c)$ so α is the unique solution i $(0,1)$ to

$$1 - \alpha = e^{-c\alpha}$$

So now we can conclude that if we let Y be the number of small vertices we have $\mathbb{E}\,Y = n(1-\alpha+o(1))$ as each of n vertices has probability $(1-\alpha+o(1))$ of being small. We note that this implies $\mathbb{E}\,Y \to \infty$ as $n \to \infty$.

Using indicator functions for pairs of vertices being small we get:

$$\mathbb{E}(Y(Y-1)) = \sum_{i \neq j} \mathbb{E}\, I_{v_i, v_j \text{ are both small}} =$$

$$= n \sum_{j=2}^{n} \mathbb{P}(v_1, v_j \text{ are both small}) \leq n\rho(n,p)(k_- \cdot 1 + n\rho(n - O(k_-), p))$$

Where we have split the sum on v_j being in the same component as v_1 (there being almost surely at most k_- such vertices) and such vertex has probability 1 of being small, while any v_j not in the same component as v_1 has probability $\rho(n - O(k_-), p)$ of being in the small component as size of v_1's component is $O(k_-)$ and there are at most n such v_j.

$$\mathbb{E}(Y(Y-1)) \leq n\rho(n,p)(k_- + n\rho(n - O(k_-), p)) = \mathbb{E}\,Y \cdot (\mathbb{E}\,Y(o(1)+1)) = (\mathbb{E}\,Y)^2(1+o(1))$$

Now using Chebyshev's Inequality we get:

$$\mathbb{P}(|Y - \mathbb{E}\,Y| \geq t) \leq \frac{\operatorname{Var}(Y)}{t^2} = \frac{\mathbb{E}(Y^2 - Y) + \mathbb{E}\,Y - (\mathbb{E}\,Y)^2}{t^2} \leq \frac{(\mathbb{E}\,Y)^2 o(1) + \mathbb{E}\,Y}{t^2}$$

Given $\epsilon > 0$ by putting $t = \epsilon\,\mathbb{E}\,Y$ we get:

$$\mathbb{P}(Y \geq (1+\epsilon)\,\mathbb{E}\,Y) \leq \mathbb{P}(|Y - \mathbb{E}\,Y| \geq \mathbb{E}\,Y) \leq \frac{o(1)}{\epsilon^2} + \frac{1}{\mathbb{E}\,Y \epsilon^2} \to 0$$

So almost certainly there are at most $(1+\epsilon)(1-\alpha+o(1))n$ small vertices as claimed.

□

Therefore, the giant component will almost certainly have at least $n - (1+\epsilon)n(1-\alpha+o(1)) = n\alpha\left(1 - o(1) - \epsilon\left(\frac{1}{\alpha} - 1 + o(1)\right)\right)$ vertices. Noting that for n large enough we have $o(1) + \epsilon\left(\frac{1}{\alpha} - 1 + o(1)\right) \leq 2\epsilon$.

Therefore as claimed size of the largest component is at least $n\alpha(1 + o_p(1))$.

This completes the proof of Theorem 4.

□

There are many further deep and beautiful results concerning the component structure of random evolution of graphs some of which we have mentioned in the intuitive description we gave in the beginning of this section they are all proven in either [3] or [5] or in both.

5 References.

[1] B. Bollobás, Modern Graph Theory, 1st ed., Springer, (1998).

[2] J. Komlós and E. Szemeredi, Limit distributions for the existence of Hamilton circuits in a random graph, Discrete Mathematics 43 (1983), 55 − 63.

[3] S. Janson, T. Łuczak and A. Ruciński, Random Graphs, Wiley, New York, (2000).

[4] P. Erdős and A. Rényi, On the strength of connectedness of a random graph, Acta Math. Acad. Sci. Hung., 12, (1961), pp. 261 − 267.

[5] B. Bollobás, Random graphs, 2nd ed., Cambridge University Press, (2001).

[6] M. Krivelevich, Long paths and Hamiltonicity in random graphs, Lectures on Random Graphs, Geometry and Asymptotic Structure, (2013),

http://web.mat.bham.ac.uk/combinatorics/LMS-EPSRC/index.html.