1. Components Sizes for $G_{n,p}$

We discuss the size of the components in $G_{n,p}$ for p bounded away from $\frac{1}{n}$. Our main result, proven by Erdos and Renyi in 1960, is that for any $\epsilon > 0$ there is a $\delta > 0$ such that (i) if $p < \frac{1-\epsilon}{n}$ then almost surely the largest component of $G_{n,p}$ has size $O(log\ n)$ while if $p > \frac{1+\epsilon}{n}$ then almost surely $G_{n,p}$ has a component of size at least δn, and all other components have size $\mathcal{O}(log\ n)$. We also discuss the probability that $G_{n,p}$ is connected for various values of p.

We use a probabilistic approach which we will generalize and apply to random regular graphs on a fixed degree sequence. Our focus is on presenting simple arguments. We do not attempt to obtain the strongest possible results, nor will we mention much about the history of the subject. We also do not attempt to treat $p = \frac{1+o(1)}{n}$. See for an extensive study of the situation there. That paper considers a stochastic process where we add the edges to G one at a time, and study the random graph $G_{n,m}$ which is chosen uniformly from those with n vertices and m edges. This allows for much more precise results but we will not need such precision.

We let Y_s be the number of components of $G_{n,p}$ which have s vertices. Every such component contains at least one tree which spans s vertices such that no edges of $G_{n,p}$ links the tree to the remaining vertices. So, the expected number of such trees is an upper bound on $E(Y_S)$.

Since there are $\binom{n}{s}$ ways of choosing the vertices of such a tree, and s^{s-2} ways of choosing the $s-1$ edges of a tree on a chosen set of vertices, this yields:

$$E(Y_s) \leqslant \binom{n}{s} s^{s-2} p^{s-1} (1-p)^{s(n-s)}.$$

Since $s! \geqslant \frac{s^s}{e^s}$, we have

$$E(Y_s) \leqslant \frac{(epn(1-p)^{n-s})^s}{ps^2}.$$

Now, letting Z_s be the number of components with S vertices which induce trees, a similar computation shows:

$$E(Y_s) \geqslant E(Z_s) \geqslant \binom{n}{s} s^{s-2} p^{s-1} (1-p)^{s(n-s)+\frac{(s-2)(s-1)}{2}}.$$

Very soon, we will restrict our attention to s which are $O(logn)$ and p which are $O(\frac{logn}{n})$. In this situation our two bounds determine $E(Y_s)$ to within a $1 + o(1)$ factor.

1.1. **The Connectivity Threshold.**

As discussed in the background notes for the Markov Chain course, we can couple $G_{n,p}$ and $G_{n,p'}$ for $p < p'$ by assigning a uniform [0..1] variable to each edge, and putting an edge into $G_{n,p}$ (respectively p') if this real is less than p (resp p'). This couples any choice of $G_{n,p}$ with a choice of $G_{n,p'}$ of which it is a subgraph. Hence the probability that $G_{n,p}$ is connected cannot decrease when p increases. Furthermore, the expected value of Y_1 strictly decreases as p increases.

We note that a graph is disconnected precisely if it has a component which contains at most half of its vertices. So, $P(G_{n,p}\ is\ not\ connected) = P(\Sigma_{s=1}^{\lfloor n/2 \rfloor} Y_s > 0)$.

Consider now any positive f going to infinity with n with $f(n) < log\ n$ and $p = \frac{log\ n + f(n)}{n}$. Then $E(Y_1)$ is $o(1)$ and $E(Y_2)$ is $o(\frac{log\ n}{n})$. Moreover for s between 3 and $\frac{n}{2}$ and any p with $\frac{log\ n}{n} \leqslant p \leqslant \frac{2log\ n}{n}$, we have $(1-p)^{n-s} \leqslant \sqrt{n}$ thus $E(Y_s) \leqslant \frac{(2elog\ n/\sqrt{n})^s}{ps^2}$.

So, for any positive f which goes to to infinity with n and satisfies $f(n) < log\ n$ and corresponding $p = \frac{logn + f(n)}{n}$, we have

$$E(\Sigma_{i=1}^{\frac{n}{2}} Y_s) \leqslant o(1) + O(\frac{logn}{n}) + \sum_{s=3}^{\lfloor \frac{n}{2} \rfloor} \frac{(\frac{2elog\ n}{\sqrt{n}})^s}{ps^2} = o(1).$$

Thus applying Markov's Inequality for any such f and p we know that $P(G_{n,p}\ is\ connected) = 1 - o(1)$. Since the probability $G_{n,p}$ is connected is nondecreasing this is true for any p such that $pn - logn = \omega(1)$.

On the other hand, for any positive f going to infinity with n and $p = \frac{logn - f(n)}{n}$, $E(Y_1) = n(1-p)^{n-1} = \omega(1)$. Further,

$$E({Y_1}^2) = n(1-p)^{n-1} + n(n-1)(1-p)^{2n-3} \leqslant E(Y_1) + (1-p)^{-1}E(Y_1)^2.$$

So, $E(Y_1^2) - E(Y_1)^2 = o(1)E(Y_1)^2$, and applying Chebbyshev, we obtain:

$P(Y_1 = 0) \leqslant P(|Y_1 - E(Y_1)| \geqslant E(Y_1)) = o(1)$.

1.2. **The Existence of A Giant Component.**

To begin we show;

Theorem 1.1. *For all* $\epsilon > 0$ $\exists A_\epsilon, b_\epsilon > 0$ *such that for all* p *with* $|\frac{p}{n} - 1| > \epsilon$,

$$P(\sum_{A_\epsilon logn < s < b_\epsilon n} Y_s > 0) = o(1).$$

Proof. We can assume that for $d = pn$, we have $d \leqslant \frac{2logn}{n}$ as we showed in the last section that otherwise $G_{n,p}$ is connected with probability 1-o(1). I.e. the desired result is true even if we sum over s from 1 to $\frac{n}{2}$. Furthermore, we can assume that $p \geqslant n^{-3}$ as otherwise with probability $1 - o(1)$ there are no edges and hence $Y_s = 0$ for all $s > 1$.

As noted above,

$$E(Y_s) \leqslant \frac{(epn(1-p)^{n-s})^s}{ps^2}. \leqslant n^3(epn(1-p)^{n-s})^s$$

Since $d \leqslant 2log\ n$, for $s \leqslant \frac{n}{2}$ this is $O(n^3((1+o(1))ede^{-d/2})^s)$. So we need only consider $d \leqslant 100$ as otherwise $E(Y_s) \leqslant n^3 10^{-s}$ and we are done by applying Markov's Inequality.

Now the derivative f of $p(1-p)^{n-1}$ is $(1-p)^{n-1} - p(n-1)(1-p)^{n-2} = (1-p)^{n-2}(1-pn)$. The only zero of f between 0 and 1 is at $p = \frac{1}{n}$ which is a maximum of $p(1-p)^{n-1}$ in this range. Furthermore $(1-\frac{1}{n})^{n-1}$ lies between e^{-1} and $(1+\frac{2}{n-1})e^{-1}$, so at $p = \frac{1}{n}$, $(epn(1-p)^{n-1})$ lies between 1 and $1+\frac{2}{n-1}$. Finally letting $\epsilon' = min(\epsilon, \frac{1}{2})$, we see that for x with $|xn-1|$ lying between $\frac{\epsilon'}{2}$ and ϵ' and large n, $|f(x)|$ is at least $\frac{\epsilon'}{2e^2 n}$. Integrating f we see that for p with $|pn-1| \geqslant \epsilon \geqslant \epsilon'$:

$$epn(1-p)^{n-1} \leqslant (1+\frac{2}{n-1} - \frac{(\epsilon')^2}{4e^2}) \leqslant 1 - \frac{(\epsilon')^2}{5e^2}.$$

Now, since $d \leqslant 100$, if we take $b_\epsilon > 0$ sufficiently small in terms of ϵ for $s \leqslant b_\epsilon n$, we can ensure $(1-p)^{s-1} \geqslant (1-\frac{100}{n})^{b_\epsilon n} \geqslant \sqrt{(1-\frac{(\epsilon')^2}{5e^2})}$. So for such s, $E(Y_s) \leqslant n^3(1-\frac{(\epsilon')^2}{5e^2})^{-s/2}$. So we can choose A_e for which the theorem holds. □

We now show(slightly differently than in the powerpoints):

Theorem 1.2. *For all $\epsilon > 0$ there is a $b_\epsilon > 0$ and an A_ϵ such that:*

(i) If $pn - 1 > \epsilon$ then with probability $1 - o(1)$, $G_{n,p}$ has a component of size exceeding $b_\epsilon n$,
(ii) If $pn - 1 < \epsilon$ then with probability $1 - o(1)$ every component of $G_{n,p}$ has size at most $A_e log\ n$.

Proof. We note that Theorem 1.1 implies that to prove (i) it is enough to show that if $pn - 1 > \epsilon$ then with probability $1 - o(1)$, $G_{n,p}$ has a component of size exceeding $A_e log\ n$ and to prove (ii) it is enough to show that for $pn - 1 < \epsilon$ the probability $G_{n,p}$ has a component of size exceeding $b_e n = o(1)$.

To do so we consider an iterative exploration process which runs for $i = \lceil \sqrt{n} - log^2 n \rceil$ steps. where at the start of iteration i we have a set O_i of discovered but as yet unexplored vertices and a set E_i of explored vertices. We intialize $O_1 = v_1$ and $E_1 = \emptyset$.

In iteration i, we take the lowest indexed vertex v in O_i, we expose the edges from it to $V - E_i - O_i$. We set $E_{i+1} = E_i + v$ and $O_{i+1} = O_i - v \cup (N(v) \cap (V - E_i - O_i))$.

If $O_i \geqslant (logn)^2$ we terminate.

(*) If $O_i = 0$ we add the lowest indexed vertex not in E_i to O_i.

Ignoring the vertices added in (*), the number of vertices added to O_i in iteration i is $Bin(n - |O_i| - |E_i|, p)$. Now, $n - |O_i| - |E_i|$ lies between $n - \sqrt{n} - 1$ and n.

If the component containing v_1 has at least $\sqrt{n}$ vertices then we never added vertices in (*), and we had to add at least $\sqrt{n} - 1 - (log\ n)^2$ vertices to O_i. The probability this occurs is at most the probability that $Bin(n' = \lceil \sqrt{n} - (log\ n)^2 \rceil n, p) > \sqrt{n} - (log\ n)^2 - 1$. For $pn < 1 - \epsilon$ the expectation of $Bin(n', p) < (1-\epsilon)\sqrt{n}$ and an application of the Chernoff bound shows that for large n, the probability that v_1 is in a component with greater than $\sqrt{n}$ vertices is less than $e^{\frac{\epsilon^2 \sqrt{n}}{12}} = o(\frac{1}{n})$. By symmetry each v_i has the same probability of being in a component of size exceeding $\sqrt{n}$, so the expected number of such vertices is $o(1)$ and almost surely no such component exists.

Now, if $G_{n,p}$ has no component of size greater than $(log\ n)^2$ then we never terminate, and the the total number of vertices added to O_i (not counting those added in (*)) is at most $\sqrt{n} + 1$. This is at most the probability that $Bin(n'' = \lceil \sqrt{n} - (log\ n)^2 \rceil (n - \sqrt{n} - 1), p) \leqslant \sqrt{n} + 1$. For $pn > 1 + \epsilon$ and large n, the expectation of $Bin(n'', p) > (1 + \frac{\epsilon}{2})\sqrt{n}$. An application of Chernoff's Bounds shows that the probability $Bin(n'', p) \leqslant \sqrt{n} + 1 = o(1)$. Hence with probability $1 - o(1)$, $G_{n,p}$ has a component of size exceeding $(log\ n)^2$.

□

2. Some Results on Random Regular graphs

2.1. **Switchings.**

By a switching in a graph or multigraph G, we mean the deletion of two disjoint edges xy and wv, and the addition of the edges xw and yv. We note that given two disjoint edges in a graph there are two possible switchings using these edges.

Our interest in this operation is that it does not change the degree sequence of the graph. We will use it to bound the probability of specific events with respect to the multigraph $H_{\mathcal{D}}$ created by the configuration model given a degree sequence $\mathcal{D}$, and the random graph $G_{\mathcal{D}}$ chosen uniformly from those with this degree sequence.

We note that if $\mathcal{A}$ and $\mathcal{B}$ are two families of graphs and for every graph in $\mathcal{A}$ there are at most Δ switching which yield a graph in $\mathcal{B}$ and for every graph in $\mathcal{B}$ there are at least δ switchings which yield a graph in $\mathcal{A}$ then we have:

$$\delta|\mathcal{B}| \leqslant \#\ of\ switchings\ between\ \mathcal{A}\ and\ \mathcal{B} \leqslant \Delta|\mathcal{A}|.$$

Hence $|\mathcal{B}| \leqslant \frac{\Delta}{\delta}|\mathcal{A}|$. So, if $\mathcal{A}$ and $\mathcal{B}$ are families of graphs with degree distribution $\mathcal{D}$ then $P(G_{\mathcal{D}} \in \mathcal{B}) \leqslant \frac{\Delta}{\delta} P(G_{\mathcal{D}} \in \mathcal{A})$.

In the same vein if $\mathcal{A}$ and $\mathcal{B}$ are families of matchings on the vertex set of $M_{\mathcal{D}}$ then $P(M_{\mathcal{D}} \in \mathcal{B}) \leqslant \frac{\Delta}{\delta} P(M_{\mathcal{D}} \in \mathcal{A})$.

2.2. **The Largest Component of a Uniform 2-Factor.**

We let $\mathcal{D}$ be the degree sequence consisting of n twos. Then $G_{\mathcal{D}}$ is a uniformly chosen *2-factor* from those with n vertices.

Now for any constant ϵ 0 and 1, the expected number of cycles of length at least ϵn in $G_{\mathcal{D}}$ is

$$\Sigma_{i=\lceil \epsilon n \rceil}^{n} \frac{n!}{2i((n-i)!)} \frac{2^{i-1}}{\Pi_{j=1}^{i} 2n-2j+1} \geqslant \Sigma_{i=\lceil \epsilon n \rceil}^{n} \frac{1}{4i} \approxeq \frac{ln \frac{1}{\epsilon}}{4}.$$

Since a 2 factor with n vertices can have at most $\frac{1}{\epsilon}$ cycles which have length at least ϵn, it follows that the probability $G_{\mathcal{D}}$ has a cycle of length at least ϵn is essentially $\geqslant \epsilon \frac{ln \frac{1}{\epsilon}}{4}$.

We now upper bound the probability that $G_{\mathcal{D}}$ has a cycle of length at least ϵn. To do so we let A_i (resp. B_i) be the set of 2-factors on n vertices which have at least one (resp. no) cycle of length at least ϵn and i cycles of length strictly between $\frac{\epsilon n}{3}$ and ϵn.

We note that A_i and B_i are empty for $i > \lfloor \frac{3}{\epsilon} \rfloor$. We claim that for large n $\sum_{i=1}^{\lfloor \frac{3}{\epsilon} \rfloor} P(B_i) \geqslant \frac{(4\epsilon^{-2}+1)^{\frac{-\epsilon}{3}}}{\frac{3}{\epsilon}+2}$, which is a lower bound on the probability that a uniform random 2-factor has no cycle of length at least ϵn.

To prove our claim we assume the contrary and obtain a contradiction. We note that for every 2-factor in A_i, if we swap two edges of its longest cycle which are at distance exceeding $\frac{\epsilon n}{3}$ then we obtain a two factor in $A_{i+1} \cup B_{i+1}$. For large n, there are at least $\frac{\epsilon^2 n^2}{4}$ switches involving such a pair of edges (each pair yields two switches). There are at most $2\binom{n}{2}$ switches from a 2-factor in $A_{i+1} \cup B_{i+1}$ to a 2-factor in A_i. Thus, we have $|A_i| \leqslant 4\epsilon^{-2}(|A_{i+1}| + |B_{i+1}|)$. Inductively, we have that for j from $\lfloor \frac{3}{\epsilon} \rfloor$ to 0:

$$P(A_j \cup B_j) \leqslant (4\epsilon^{-2}+1)^{\lfloor \frac{3}{\epsilon} \rfloor - j} \frac{(4\epsilon^{-2}+1)^{\frac{-\epsilon}{3}}}{\frac{3}{\epsilon}+2} \leqslant \frac{1}{\frac{3}{\epsilon}+2}$$

.

But this contradicts the fact that $\sum_{j=0}^{\lfloor \frac{3}{\epsilon} \rfloor} P(A_j \cup B_j) = 1$

2.3. **The Probability That $H_{\mathcal{D}}$ is simple.**

For a fixed integer d, we consider the degree sequence $\mathcal{D}$ where all degrees are d, a random matching $M_{\mathcal{D}}$ and corresponding multigraph $H_{\mathcal{D}}$.

The probablity two specific copies of a vertex v are joined by an edge of $M_{\mathcal{D}}$ is $\frac{1}{dn-1}$, so the expected number of loops in $H_{\mathcal{D}}$ is $\frac{\binom{d}{2}n}{dn-1} \leqslant \frac{d}{2}$.

In the same vein, for every copy of a vertex, given a choice for the other endpoint of the edge of $M_{\mathcal{D}}$ containing it which does not create a loop, the probability that this edge is parallel to some other edge is at most $\frac{(d-1)^2}{dn-3}$. Hence the expected number of edges parallel to another edge is at most $\frac{dn(d-1)^2}{2(dn-3)}$ which is less than $\frac{d^2-d}{2}$ for large n.

Finally the expected number of triples of edges all joining the same two points is $\frac{\binom{n}{2}\binom{d}{3}^2 3!}{(dn-1)(dn-3)(dn-5)} = o(1)$.

Letting $l(H)$ be the number of loops in H. we have that $E(l(H_{\mathcal{D}})) \leqslant \frac{d}{2}$ So Markov's Inequality tells us that $P(l(H_{\mathcal{D}}) > 2d) \leqslant \frac{1}{4}$. Letting $p(H)$ be the the number of parallel edges in H, we have that $E(p(H_{\mathcal{D}})) \leqslant \frac{d^2-d}{2}$ so $P(p(H_{\mathcal{D}}) > 2d^2 - 2d) \leqslant \frac{1}{4}$.

So letting $A_{j,r}$ be the set of those matchings on $V(M_{\mathcal{D}})$ corresponding to a choice of $H_{\mathcal{D}}$ with $l(H_{\mathcal{D}}) = j, p(H_{\mathcal{D}}) = r$ and such that $H_{\mathcal{D}}$ contains no triple of parallel edges, we know that for large n,

$$\Sigma_{j=0}^{2d}\Sigma_{k=0}^{d^2-d}P(G_{\mathcal{D}}\ in\ A_{j,2k}) \geqslant \frac{1}{3}.$$

We note that to switch from $A_{j,2k}$ to $A_{j-1,2k}$, we can pick any of the l loops as one of the edges and then pick any of the at least $\frac{nd}{2} - (d-2)d - j - 2k$ edges which are not a loop, not a parallel edge, and not incident to any neighbour of the vertex which is the endpoint of this loop. To switch from $A_{j-1,2k}$ to $A_{j,2k}$ the two edges we switch on must have a common endpoint so we have at most $2n\binom{d}{2}$ choices. It follows that for $k \leqslant d^2 - d$, $j \leqslant 2d$ and n large, $|A_{j-1,2k}| \geqslant \frac{j(\frac{nd}{2}-(d-2)d-2d^2)}{nd(d-1))}|A_{j,2k}| \geqslant \frac{j}{2d}|A_{j,2k}|$. Thus, inductively, $|A_{0,2k}| \geqslant |A_{j,2k}|\frac{1}{2d^j}$ for all such j and k, This implies that $|A_{0,2k}| \geqslant \frac{1}{(d+1)2d^{2d}}\sum_{j=0}^{2d}|A_{j,2k}|$. Hence $\sum_{k=1}^{d^2-d}P(M_{\mathcal{D}} \in A_{0,2k}) \geqslant \frac{1}{6(d+1)d^{2d}}$.

Now, to switch from $A_{0,2k}$ to $A_{0,2k-2}$ we can pick any of the $2k$ parallel edges and any of the at least $\frac{nd}{2} - 2(d-1)d - 2k$ edges which are not parallel with any other edge, and do not have an endpoint incident to a neighbour of the first edge picked. To switch from $A_{0,2k-2}$ to $A_{0.2k}$ we must pick two edges, two of whose endpoints are joined by a third edge. There are at most $\frac{dn}{2}d(d-1)$ such pairs. It follows that for k between 1 and $d^2 - d$ and large n, $|A_{0,2k-2}| \geqslant \frac{1}{d^2}|A_{0,2k}|$. It follows that $|A_{0,0}| \geqslant \frac{1}{(d^2-d+1)d^{2d^2-2d}}\sum_{k=0}^{d^2-d}|A_{0,2k}|$. Hence,

$$P(H_{\mathcal{D}}\ simple) = P(M_{\mathcal{D}}\ in\ \mathcal{A}_{0,0}) \geqslant \frac{1}{6(d+1)d^{2d}}\frac{1}{(d^2-d+1)d^{2d^2-2d}} = \frac{1}{6(d+1)(d^2-d+1)d^{2d^2}}.$$

3. Exercises

(1) Consider the degree sequence $\mathcal{D}$ all of whose entries are d. Show that if $d \geqslant 5$ then the expected number of partitions of V into two sets with no edges between them is $o(1)$ and hence almost surely $G_{\mathcal{D}}$ is connected. Harder: do this for $d = 4$. Hardest: do this for $d = 3$.

(2) Consider the degree sequence $\mathcal{D}$ which has $n = 2k$ elements, k of which are some fixed $d > 1$ and k of which are 1. Show that the probability that $G_{\mathcal{D}}$ is connected is o(1).

(3) Show that for $d \geqslant 10$, there is an $A_d > 0$ and a $b_d > 0$ such that the probability that the $G_{\mathcal{D}}$ of question (2) contains a component of size between $A_d log\ n$ and $b_d n$ is o(1). Hint: you will need to split the proof into two parts. You should start by showing the probability there is such a component for which the proportion of vertices of degree d exceeds $\frac{1+\epsilon}{d-1}$ for some $\epsilon > 0$ is o(1).

(4) Show that for $d \geqslant 10$, there is a $c_d n$ such that the probability the $G_{\mathcal{D}}$ of Question 2 contains a component of size exceeding $c_d n$ is 1-o(1).

4. Further Reading

https://www.math.uwaterloo.ca/ nwormald/papers/regsurvey.pdf

https://arxiv.org/pdf/math/9310236.pdf