2.4 The giant component of the Erdős–Rényi random graph

2.4.1 Non-rigorous discussion

We know that if $pn = o(1)$ then there are no triangles. In a similar manner it can be shown that there are no cycles of any order in $\mathcal{G}(n,p)$. This means that most components of the random graph are trees and isolated vertices. For $p > c\log n/n$ for $c \geq 1$ the random graph is connected a.a.s. What happens in between these stages? It turns our that a unique giant component appears when $p = c/n$, $c > 1$. We first study the appearance of this largest component in a heuristic manner. We define *the giant component* as the component of $\mathcal{G}(n,p)$, whose order is $\mathcal{O}(n)$.

Let u be the frequency of the vertices that do not belong to the giant component. In other words, u gives the probability that a randomly picked vertex does not belong to the giant component. Let us calculate this probability in a different way. Pick any other vertex. It can be either in the giant component or not. For the original node not to belong to the giant component these two either should not be connected (probability $1-p$) or be connected, but the latter vertex is not in the giant component (probability pu). There are $n-1$ vertices to check, hence

$$u = (1 - p + pu)^{n-1}.$$

Recall that we are dealing with $pn \sim \lambda$, hence it is convenient to use the parametrization

$$p = \frac{\lambda}{n}, \quad \lambda > 0.$$

Note that λ is the mean degree of $\mathcal{G}(n,p)$. We get

$$\begin{aligned}
u = \left(1 - \frac{\lambda}{n}(1-u)\right)^{n-1} &\implies \\
\log u(n-1)\log\left(1 - \frac{\lambda}{n}(1-u)\right) &\implies \\
\log u \approx -\frac{\lambda(n-1)}{n}(1-u) &\implies \\
\log u \approx -\lambda(1-u) &\implies \\
u = e^{-\lambda(1-u)}, &
\end{aligned}$$

where the fact that $\log(1+x) \approx x$ for small x was used.

Finally, for the frequency of the vertices in the giant component $v = 1-u$ we obtain

$$1 - v = e^{-\lambda v}.$$

This equation always has the solution $v = 0$. However, this is not the only solution for all possible λs. To see this consider two curves, defined by $f(v) = 1 - e^{-\lambda v}$ and $g(v) = v$. Their intersections give the roots to the original equation. Note that, as expected, $f(0) = g(0) = 0$. Note also that $f'(v) = \lambda e^{-\lambda v} > 0$ and $f''(v) = -\lambda^2 e^{-\lambda v} < 0$. Hence the derivative for $v \leq 0$ cannot be bigger than at $v = 0$, which is $f'(0) = \lambda$. Hence (see the figure), if $\lambda \leq 0$ then there is unique trivial solution $v = 0$, but if $\lambda > 1$ then another positive solution $0 < v < 1$ appears.

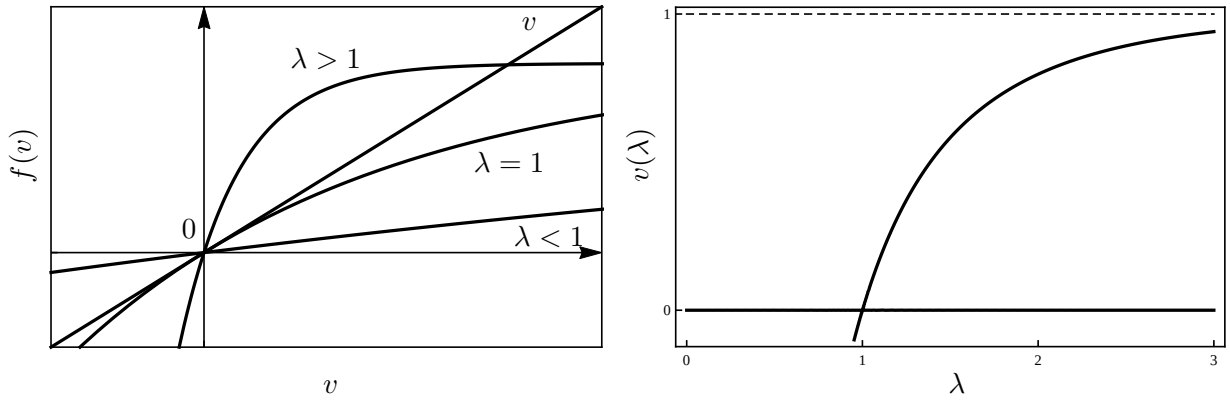

Figure 2.1: The analysis of the number of solutions of the equation $1 - v = e^{-\lambda v}$. The left panel shows that if $v \geq 0$ then it is possible to have one trivial solution $v = 0$ in the case $\lambda \leq 1$, and two solutions if $\lambda > 1$. The right panel shows the solutions (thick curves) as the functions of λ

Technically, we only showed that if $\lambda \leq 1$ then there is no giant component. If $\lambda > 1$ then we use the following argument. Since λ gives the mean degree, then, starting from a randomly picked vertex, it will have λ neighbors on average. Its neighbors will have λ^2 neighbors (there are λ of them and each has λ adjacent edges) and so on. After s steps we would have λ^s vertices within the distance s from the initial vertex. If $\lambda > 1$ this number will grow exponentially and hence most of the nodes have to be connected into the giant component. Their frequency can be found as the nonzero solution to $1 - v = e^{-\lambda v}$. Of course this type of argument is extremely rough, but the fact is that it can be made absolutely rigorous within the framework of the branching processes.

Moreover, the last heuristic reasoning can be used to estimate the diameter of $\mathcal{G}(n, p)$. Obviously, the process of adding new vertices cannot continue infinitely, it has to stop when we reach all n vertices:

$$\lambda^s = n.$$

From where we have that

$$s = \frac{\log n}{\log \lambda}$$

approximates the diameter of the Erdős–Rényi random graph. It is quite surprising that the exact results are basically the same: It can be proved that for $np > 1$ and $np < c \log n$, the diameter of the random graph (understood as the diameter of the largest connected component) is concentrated on at most four values around $\log n / \log np$.

Finally we note that since the appearance of the giant component shows this threshold behavior (if $\lambda < 1$ there is no giant component a.a.s., if $\lambda > 1$ the giant component is present a.a.s.) one often speaks of a *phase transition*.

2.4.2 Rigorous results for the appearance of the giant component

```
Add the discussion on the branching processes
and relation to the appearance of the giant component.
```

2.5 The evolution of the Erdős–Rényi random graph

Here I summarize the distinct stages of the evolution of the Erdős–Rényi random graph.

Stage I: $p = o(1/n)$

The random graph $\mathcal{G}(n,p)$ is the disjoint union of trees. Actually, as you are asked to prove in one of the exam problems, there are no trees of order k if $pn^{k/(k-1)} = o(1)$. Moreover, for $p = cn^{-k/(k-1)}$ and $c > 0$, the probability distribution of the number of trees of order k tends to the Poisson distribution with parameter $\lambda = c^{k-1}k^{k-2}/k!$. If $1/(pn^{k/(k-1)}) = o(1)$ and $pkn - \log n - (k-1)\log\log n \to \infty$, then there are trees of any order a.a.s. If $1/(pn^{k/(k-1)}) = o(1)$ and $pkn - \log n - (k-1)\log\log n \sim x$ then the trees of order k distributed asymptotically by Poisson law with the parameter $\lambda = e^{-x}/(kk!)$.

Stage II: $p \sim c/n$ for $0 < c < 1$

Cycles of any given size appear. All connected components of $\mathcal{G}(n,p)$ are either trees or unicycle components (trees with one additional edge). Almost all vertices in the components which are trees ($n - o(n)$). The largest connected component is a tree and has about $\alpha^{-1}(\log n - 2.5\log\log n)$ vertices, where $\alpha = c - 1 - \log c$. The mean of the number of connected components is $n - p\binom{n}{2} + \mathcal{O}(1)$, i.e., adding a new edge decreases the number of connected components by one. The distribution of the number of cycles on k vertices is approximately a Poisson distribution with $\lambda = c^k/(2k)$.

Stage III: $p \sim 1/n + \mu/n$, the double jump

Appearance of the giant component. When $p < 1/n$ then the size of the largest component is $\mathcal{O}(\log n)$ and most of the vertices belong to the components of the size $\mathcal{O}(1)$, whereas for $p > 1/n$ the size of the unique largest component is $\mathcal{O}(n)$, the remaining components are all small, the biggest one is of the order of $\mathcal{O}(\log n)$. All the components other than the giant one are either trees or unicyclic, although the giant component has complex structure (there are cycles of any period). The natural question is how the biggest component grows so quickly. Erdős and Rényi showed that it actually happens in two steps, hence the term"double jump." If $\mu < 0$ then the largest component has the size $(\mu - \log(1+\mu))^{-1}\log n + \mathcal{O}(\log\log n)$. If $\mu = 0$ then the largest component has the size of order $n^{2/3}$, and for $\mu > 0$ the giant component has the size αn for some constant α.

Stage IV: $p \sim c/n$ where $c > 1$

Except for one giant component all the components are small, and most of them are trees. The evolution of the random graph here can be described as merging the smaller components with the giant one, one after another. The smaller the component, the larger the chance of "survival." The survival time of a tree of order k is approximately exponentially distributed with the mean value $n/(2k)$.

Stage V: $p = c\log n/n$ with $c \geq 1$

The random graph becomes connected. For $c = 1$ there are only the giant component and isolated vertices.

Stage VI: $p = \omega(n) \log n/n$ where $\omega(n) \to \infty$ as $n \to \infty$. In this range the random graph is not only connected, but also the degrees of all the vertices are asymptotically equal.

Here is the numerical illustration of the evolution of the random graph. To present it I fixed $n = 1000$ the number of vertices and generated $\binom{n}{2}$ random variables, uniformly distributed in $[0, 1]$. Hence each edge gets its only number $p_j \in [0, 1]$, $j = 1, \ldots, \binom{n}{2}$. For any fixed $p \in (0, 1)$ I draw only the edges for which $p_j \leq p$. Therefore, in this manner I can observe how the evolution of the random graph occurs for $\mathcal{G}(n, p)$.

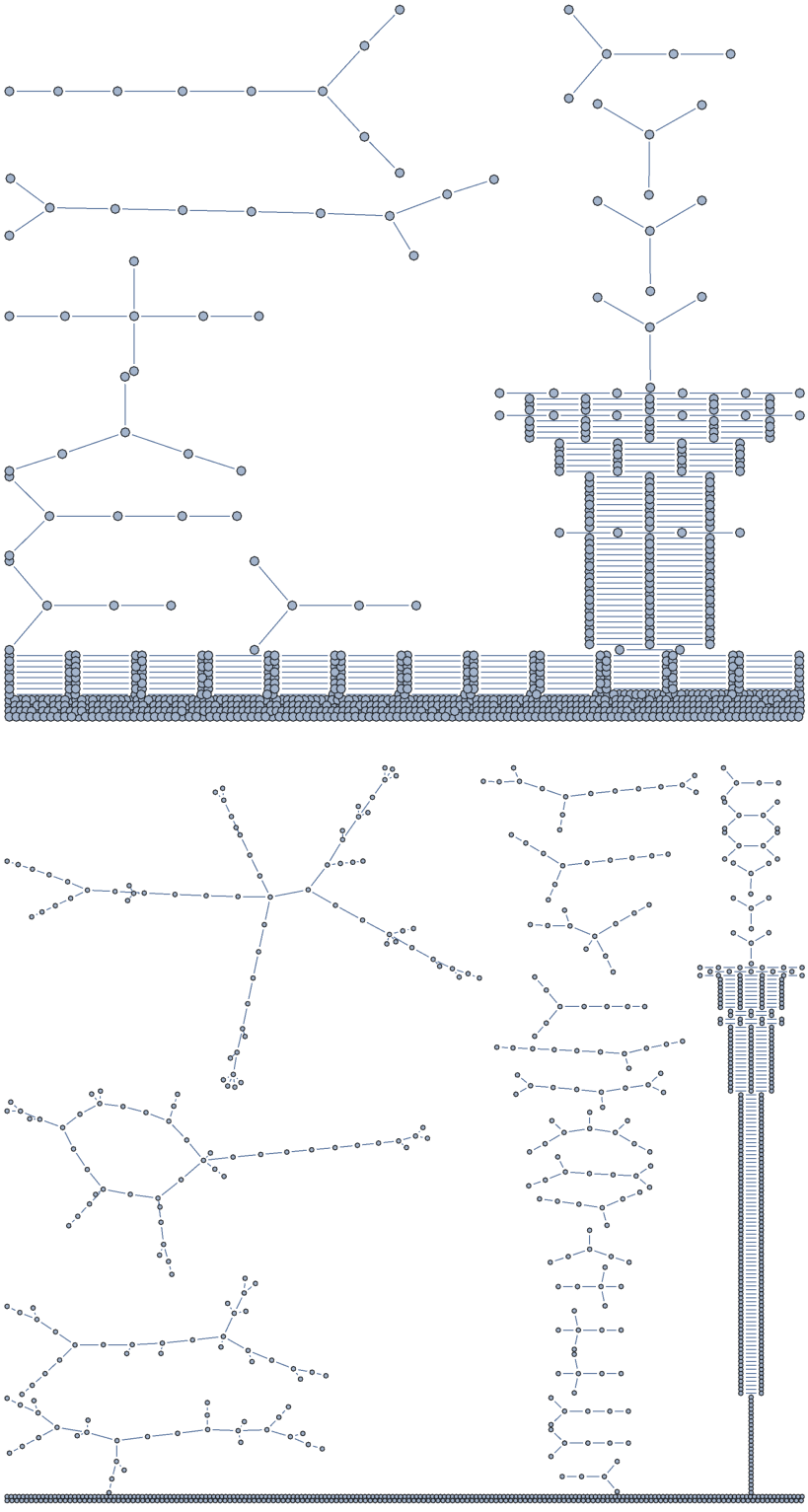

Figure 2.2: Stages 1 and 2. Graphs $G(1000, 0.0005)$ and $G(1000, 0.00095)$ are shown

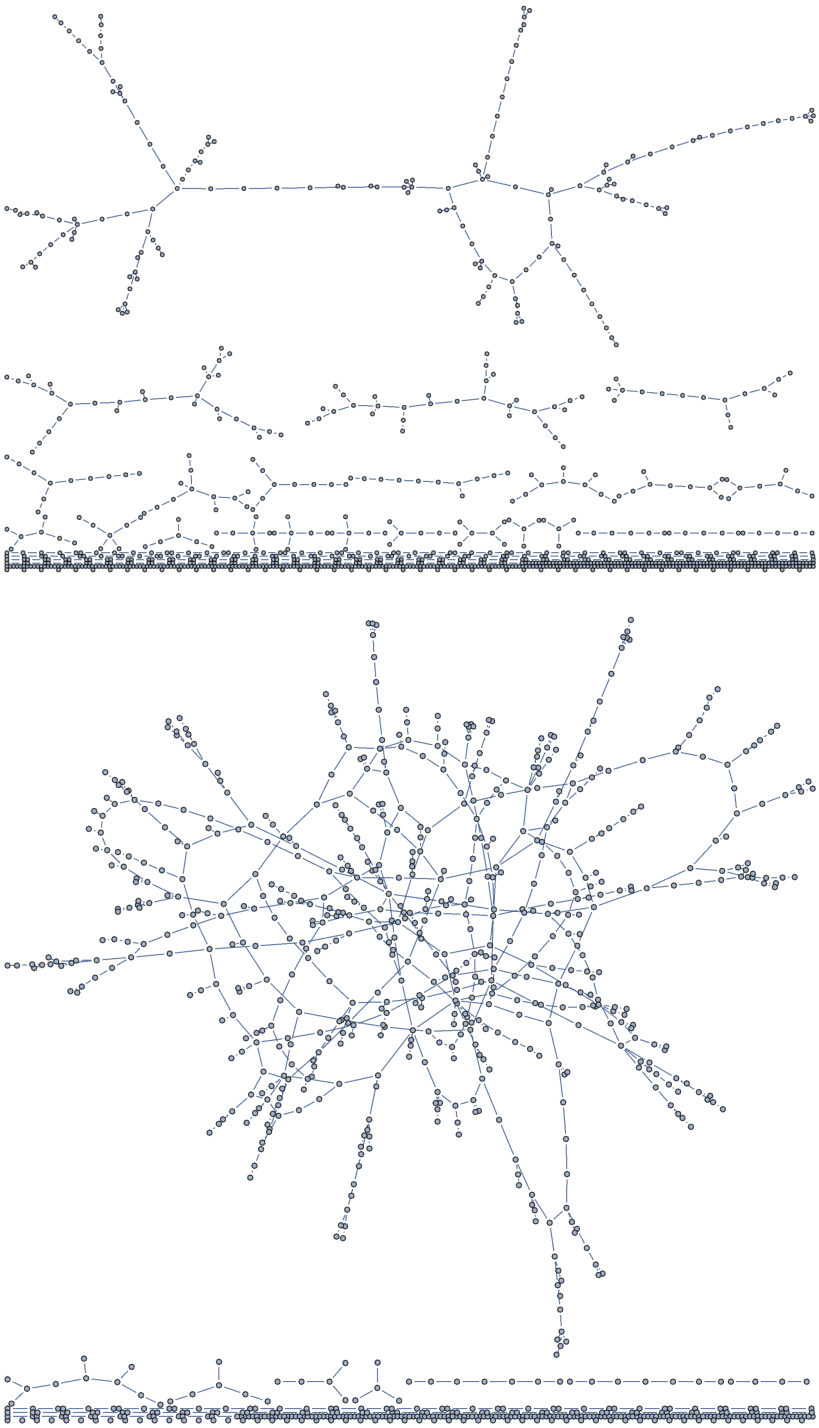

Figure 2.3: Stages 3 and 4. Graphs $G(1000, 0.001)$ and $G(1000, 0.0015)$ are shown. The giant component is born

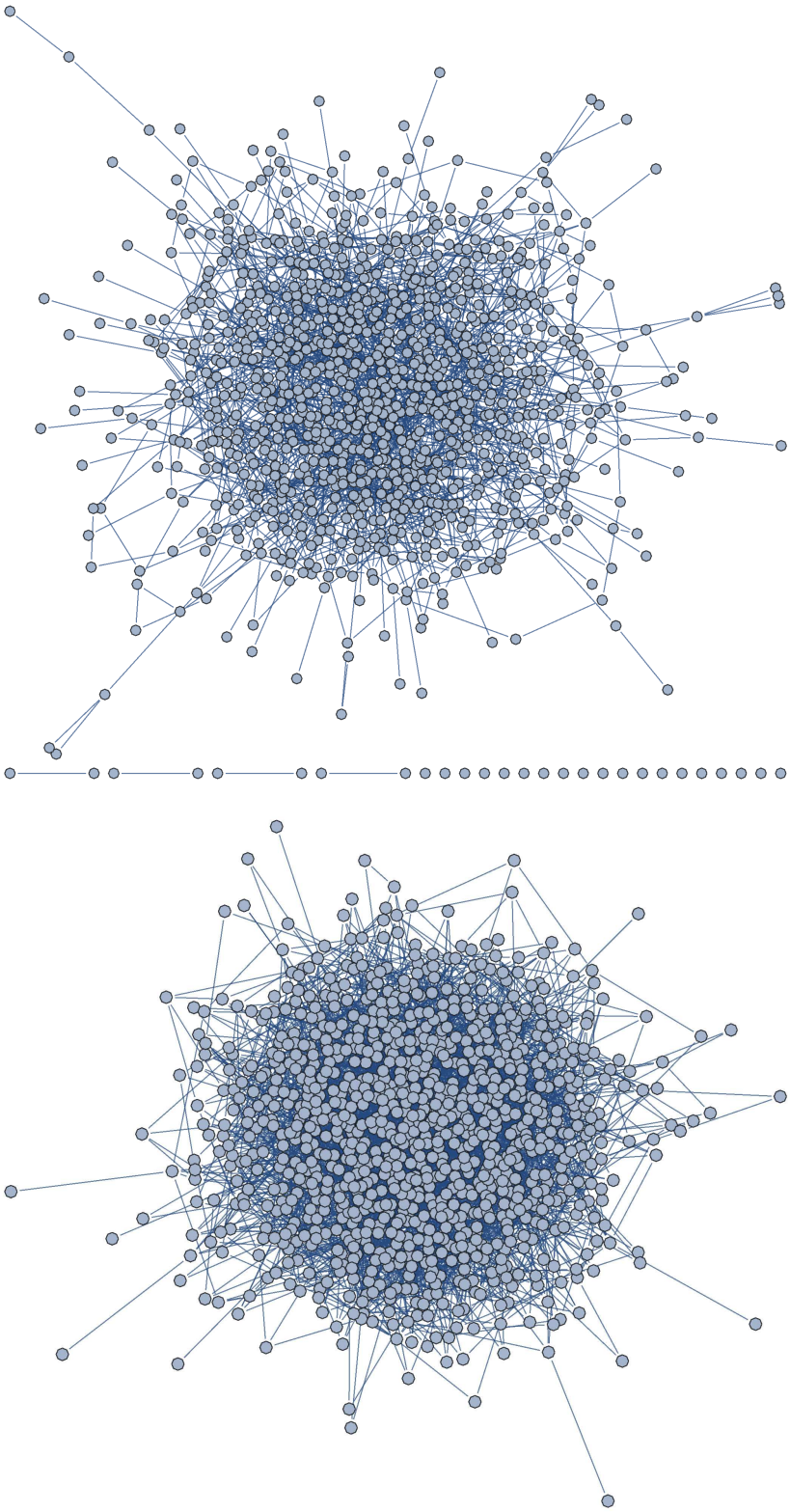

Figure 2.4: Stages 4 and 5. Graphs $G(1000, 0.004)$ and $G(1000, 0.007)$ are shown. The final graph is connected (well, almost)

Chapter 3

Generalizations of the Erdős–Renyi random graphs