Random graphs

We have so far seen a number of examples of random processes on networks, including the spread of information, competition between opinions, and random walks. In all these cases, we see that the structure of the network plays an important role in determining the long-term behaviour or final outcome, and the speed of convergence to this outcome. For example, the mixing time of a random walk is determined by the second eigenvalue of the Laplacian, which is related to the conductance of the network. The conductance also plays a role in bounding the time for information to spread on it. These findings are relevant to studying random processes on real-world networks such as the Internet or online social networks, contact networks that impact on the spread of diseases, gene regulatory networks in biology, etc. They are also relevant to designing algorithms or control mechanisms operating on such networks, e.g., routing in the Internet, targeted vaccination or quarantining for epidemics, etc.

However, many real-world networks are much too large and complex for us to expect to know their structure in full. We may only have very partial information about the structure on which to base our analysis. This is one of the motivations for studying random network models, as these can provide parsimonious representations of complex networks, while (hopefully!) capturing the key properties of interest or relevance. They can also be used to develop statistical tests of hypotheses concerning structures or patterns that arise in real-world networks.

There are a few different types of random graph models in widespread use, depending on the applications being considered. We will study the simplest of these in detail, and briefly mention some of the other models.

1 Erdős-Rényi (ER) random graphs

This random graph model was introduced by Erdős and Rényi in 1959, and has been studied extensively since then. A great deal is known about the properties of random graphs generated according to this model, and we shall look at a few of these properties. The model is parametrised by two parameters and is denoted $G(n, p)$. The notation refers to an undirected graph on n nodes, where the edge between each pair of nodes is present with probability p, independent of all other edges. This may appear to be, and indeed is, a highly over-simplified model that leaves out many features that characterise real-world networks. Nevertheless, the independence between edges makes mathematical analysis of this model especially tractable, and is one of the reasons for its continuing popularity. Another is that the model exhibits some unexpected or surprising features that may be relevant to real-world models and applications as well. One of these features is the existence of "phase transitions" for many properties where, as we change a parameter of the model, the probability of the model exhibiting that property shows a sharp transition from nearly 0 to nearly 1. We shall see some examples of this.

A closely related model, which was also studied by Erdős and Rényi in their 1959 paper introducing these models, was the $\mathcal{G}(n, M)$ model. Here, there are n nodes and M edges, and the probability distribution is uniform on the set of all such graphs. As there are only finitely many graphs with n nodes and M edges, this is a valid model. However, this description is clearly non-constructive, and it seems harder to work out properties for this model. Moreover, edges in this model are not independent; the presence of one edge makes every other edge less likely, because the total number of edges is fixed. This lack of independence also makes analysis harder. It turns out that the two models are closely related. The mean number of edges in the $G(n, p)$ model is clearly $\binom{n}{2}p$. If M is set equal to this value (rounded to a whole number), then the corresponding $\mathcal{G}(n, M)$ is 'close', at least for large n, in the sense that it assigns comparable probabilities to many events of interest; in particular, the phase transitions happen at the same parameter values.

In the following, we shall typically be interested in a sequence of random graphs indexed by n and p, where p will typically be a function of n, and n will tend to infinity. You should keep this in mind. We will be interested in understanding the limiting behaviour as n tends to infinity, rather than calculating exact probabilities for specific values of n and p. With this

justification, we are going to be extremely sloppy in many of our calculations, only keeping track of dominant terms and how they scale with n, and ignoring constants and lower order terms!

1.1 Emergence of motifs or small subgraphs

Consider an ER random graph generated according to the $G(n,p)$ model. Let H be a fixed subgraph, say a triangle. We want to ask whether a copy of this subgraph occurs anywhere within the random graph $G(n,p)$. More precisely, we would like to ask about the probability of this event. For all but the simplest subgraphs H, an exact calculation of this probability is very difficult. However it turns out that, as n tends to infinity, if we plot the probability of H appearing in $G(n,p)$ against p, this graph shows a sharp transition from a value very close to 0, to a value very close to 1. In other words, as we gradually increase p, there is a sudden change from a regime in which the random graph is extremely unlikely to contain a copy of H, to one in which it is very likely to do so (and in fact contains a large number of such copies). We shall try to characterise the value of p at which this transition takes place.

If we were to fix p at some constant value, say $3/16$, and let n tend to infinity, then the random graph $G(n,p)$ would be very likely to contain triangles. In fact, it would be very likely to contain any given fixed-size subgraph as n tended to infinity. In order to see the threshold at which triangles begin to appear in the randomg graph $G(n,p)$, we need to consider a suitable scaling for p as a function of n. It turns out that the scaling $p = n^{-\alpha}$ is the correct one. We shall look at p of this form, with α a positive constant, and ask what happens as we vary α.

For concreteness, we first continue to work with H being the triangle, and later show how to generalise the analysis to arbitrary subgraphs H. In biological applications, these small subgraphs are called motifs. Certain motifs are seen to arise often in biological networks such as gene regulation networks, and a question of interest is whether this is just chance (there are, after all, a huge number of patterns that *could* arise), or whether it might be an indicator of the functional importance of these structures. Addressing such questions requires working out how likely such structures are to arise by chance. (Obviously, the use of ER random graphs as the baseline for calculating these probabilities can be criticised.)

We now proceed with detailed calculations for the triangle. We shall start by calculating the expected number of triangles in a random graph G generated according to the distribution specified by $G(n,p)$.

Fix three nodes $i,j,k \in V$. The probability that there is a triangle on these 3 nodes is the probability that all 3 edges of this triangle are present in the random graph, which is p^3. Let $\chi_{i,j,k}$ be the indicator random variable that takes the value 1 if the triangle on these 3 nodes is present in G, and takes the value 0 otherwise. If you want to think of this formally, then the sample space is the set of all undirected graphs on a vertex set V and the probability distribution on the sample space is obtained by assigning to a graph $G=(V,E)$ the probability $p^{|E|}(1-p)^{\binom{n}{2}-|E|}$. For any fixed unordered triplet of vertices i,j,k, the random variable $\chi_{i,j,k}$ maps the sample space to $\{0,1\}$; for each graph, it specifies whether or not it contains a triangle on the vertices labelled i, j and k.

Let N_Δ be the random variable counting the number of triangles in each graph. Then,

$$N_\Delta = \sum_{i,j,k} \chi_{i,j,k}, \tag{1}$$

where the sum is taken over all unordered triplets of nodes. Clearly, there are $\binom{n}{3}$ such triplets. Hence, by the linearity of expectation,

$$\mathbb{E}[N_\Delta] = \sum_{i,j,k} \mathbb{E}[\chi_{i,j,k}] = \sum_{i,j,k} \mathbb{P}(\chi_{i,jk,}=1) = \binom{n}{3} p^3 = \frac{n^{3-3\alpha}}{3!}.$$

Thus, the expected number of triangles in a random graph drawn according to $G(n,p)$ tends to zero if $\alpha > 1$, and tends to infinity if $\alpha < 1$. There is thus a sharp change in the expected number of triangles at $\alpha = 1$.

Next, what we can say about the probability that there is a triangle in a random graph G drawn according to $G(n,p)$? If $\alpha > 1$, we see that

$$\mathbb{P}(G \text{ contains } \Delta) = \mathbb{P}(N_\Delta \geq 1) \leq \frac{\mathbb{E}[N_\Delta]}{1} \to 0 \text{ as } n \to \infty.$$

The inequality above follows from Markov's inequality for the non-negative random variable N_Δ. Thus, in this case, we can say that *with high probability*, the random graph $G(n, n^{-\alpha})$ does not contain a triangle if $\alpha > 1$. The term "with high probability" applies to an event that can be defined on a sequence of graphs, and says that the probability of this event tends to 1 as the index of the graph tends to infinity.

What happens if $\alpha < 1$? In this case, the expected number of triangles in the random graph, $\mathbb{E}[N_\Delta]$, tends to infinity. Thus, it is natural to expect that, with high probability, the random graph $G(n, n^{-\alpha})$ contains at least one triangle. While this turns out to be true for triangles, the following counter-example shows that this intuition can fail.

Counterexample We shall consider two graphs H and H' defined as follows. H' consists of a square with one of its diagonals present. Thus, it has 4 nodes and 5 edges. H is obtained from H' by adding one more node, and one more edge, say from that node to either one of the two nodes of degree 2 in H'. Thus, H has 5 nodes and 6 edges. (It might help to draw H and H' for yourself.)

We shall proceed as above to calculate the expected number of copies of H and H' in the random graph $G(n, p)$. Denote these random variables by N_H and $N_{H'}$ respectively. To estimate $\mathbb{E}[N_H]$ for example, we first choose 5 nodes, which can be done in any of $\binom{n}{5}$ ways. Earlier, there was only one way to site a triangle on 3 nodes. Now, there are many ways to site a copy of H on 5 nodes.

The exact number of ways we can do this is given by 5! divided by the size of the automorphism group of H (the number of ways we can relabel the 5 vertices such that there is an edge between the relabelled vertices if and only if there is an edge between the original ones). However, this exact number is not going to be important to us. Let us denote it by c_H, and note that c_H is a combinatorial constant that only depends on H, and not on n. As n tends to infinity, c_H will not play an important role; in particular, it will not change the threshold value of α (where $p = n^{-\alpha}$) at which $\mathbb{E}[N_H]$ jumps from nearly zero to nearly infinity.

Let us continue with the calculations. For each of the $\binom{n}{5} c_H$ ways in which H could have appeared in the random graph $G(n, p)$, the probability of each appearance is p^6, as six edges need to be present. (Note that additional edges are allowed to be present. Even if both diagonals of the square are present, for instance, we still count the graph as containing an instance of H.) Hence, by the linearity of expectation, we have

$$\mathbb{E}[N_H] = \binom{n}{5} c_H p^6 = \tilde{c}_H n^{5-6\alpha}, \tag{2}$$

where we have absorbed the constant 5! into c_H to get another constant $\tilde{c}_H$, and substituted $n^{-\alpha}$ for p. A similar calcuation for H', which has 4 nodes

and 5 edges, gives

$$\mathbb{E}[N_{H'}] = \binom{n}{4} c_{H'} p^5 = \tilde{c}_{H'} n^{4-5\alpha}. \tag{3}$$

Now, taking $\alpha = 9/11$, it is easy to see that, as n tends to infinity,

$$\mathbb{E}[N_H] = \tilde{c}_H n^{1/11} \to \infty,$$

whereas

$$\mathbb{E}[N_{H'}] = \tilde{c}_{H'} n^{-1/11} \to 0.$$

Hence, Markov's inequality tells us that the probability of the random graph $G(n,p)$ containing a copy of H' tends to zero as n tends to infinity. But H' is a subgraph of H, so if $G(n,p)$ doesn't contain H', it cannot possibly contain H. Hence, with high probability $G(n, n^{-9/11})$ contains no copy of H. Nevertheless, the expected number of copies of H it contains tends to infinity!

How is this possible? The resolution of this paradox comes from the fact that, whenever $G(n,p)$ contains a copy of H, it contains lots of copies of H. More precisely, conditional on $G(n,p)$ containing a copy of H', let us work out the expected number of copies of H it contains. Given a copy of H', it can belong to a copy of H in $2(n-4)$ ways: choose any of the $n-4$ vertices not in H', and join it to one of the two degree-2 nodes in H' with an edge. The probability that any such copy of H is present in the random graph is p, the probability that the additional edge needed to get from H' to H is present. Hence, each copy of H' 'gives rise to' approximately $2np = 2n^{2/11}$ copies of H in $G(n, n^{-9/11})$. Thus, even though the probability that there is a copy of H' in $G(n, n^{-9/11})$ scales as $n^{-1/11}$ (and the same holds for the probability that there is a copy of H), the expected number of copies of H scales as $n^{+1/11}$. □

The above counterexample tells us that it is not sufficient to look at the expected number of copies of a subgraph in order to conclude whether or not it is likely to be present in the random graph $G(n,p)$. We shall return to this question for triangles first, before extending the result to general subgraphs.

We saw above that the mean number of triangles is given by $\mathbb{E}[N_\Delta] = \binom{n}{3} p^3$, which tends to infinity as n tends to infinity if $p = n^{-\alpha}$ for $\alpha < 1$. We want to show that, in this case, the random graph contains at least one triangle with high probability. We can show this by showing that the probability

that it contains no triangles vanishes as n tends to infinity. We shall do this using Chebyshev's inequality. We have

$$\mathbb{P}(N_\Delta = 0) = \mathbb{P}(N_\Delta \leq 0) \leq \mathbb{P}(|N_\Delta - \mathbb{E}[N_\Delta]| \geq \mathbb{E}[N_\Delta]) \leq \frac{\mathrm{Var}(N_\Delta)}{(\mathbb{E}[N_\Delta])^2}. \quad (4)$$

If we can show that $\mathrm{Var}(N_\Delta)/(\mathbb{E}[N_\Delta])^2$ tends to zero as n tends to infinity, then it follows that the probability of no trianges tends to zero, and hence that the probability that there is at least one triange tends to 1, which is what we want to show.

We already know $\mathbb{E}[N_\Delta]$, so it remains to estimate the variance of the number of triangles in $G(n,p)$. To do this, let us first write the random variable N_Δ as the sum of indicator random variables:

$$N_\Delta = \sum_{A \subset V : |A| = 3} \chi_A, \quad (5)$$

where χ_A is the indicator of the event that $G(n,p)$ contains a triangle sitting on the three vertices making up the set A. This is just a rewriting of equation (1) in more convenient notation. Now, if we can express any random variable as a sum of random variables, this yields an expression for its variance. If $Y = X_1 + X_2 + \ldots + X_m$, then $\mathrm{Var}(Y) = \sum_{i,j=1}^{m} \mathrm{Cov}(X_i, X_j)$, where $\mathrm{Cov}(X_i, X_j) = \mathbb{E}[X_i X_j] - \mathbb{E}[X_i]\mathbb{E}[X_j]$ denotes the covariance of X_i and X_j, and in particular, $\mathrm{Cov}(X_i, X_i) = \mathrm{Var}(X_i)$. Thus, it follows from (5) that

$$\mathrm{Var}(N_\Delta) = \sum_{\substack{A_1, A_2 \subset V \\ |A_1|=3, |A_2|=3}} \mathrm{Cov}(\chi_{A_1}, \chi_{A_2}). \quad (6)$$

We shall now evaluate this covariance in the different scenarios possible.

If A is a set of 3 vertices, we shall write $\Delta(A)$ to denote the set of 3 edges making up the triangle on these vertices. If A_1 and A_2 are 3-vertex subsets of V, then the possible values of $|\Delta(A_1) \cap \Delta(A_2)|$ are 0, 1 or 3. In words, the vertices might be such that the triangles on them don't have any edges in common (because A_1 and A_2 have either zero or one vertices in common), have one edge in common ($|A_1 \cap A_2| = 2$) or all 3 edges in common ($A_1 = A_2$). Now, because edges in an ER random graph are mutually independent, it is clear that triangles on A_1 and A_2 are independent if these triangles have no edges in common. Hence, in this case, the indicator random variables denoting their presence are also independent, and $\mathrm{Cov}(\chi_{A_1}, \chi_{A_2}) = 0$. Next, if $|\Delta(A_1) \cap \Delta(A_2)| = 1$, then

$$\mathrm{Cov}(\chi_{A_1}, \chi_{A_2}) = \mathbb{E}[\chi_{A_1}\chi_{A_2}] - \mathbb{E}[\chi_{A_1}]\mathbb{E}[\chi_{A_2}] = p^5 - p^6. \quad (7)$$

To see the second equality note that $\chi(A_1)\chi(A_2)$ is the indicator that two (specific) triangles with one common edge are present; this corresponds to five edges being present, which has probability p^5. (The expectation of an indicator random variable is simply the probability of the indicated event.) On the other hand, each of $\chi(A_1)$ and $\chi(A_2)$ have expectation p^3.

Finally, if $A_1 = A_2$, then $\chi(A_1)\chi(A_2) = \chi(A_1)$, and we have

$$\mathrm{Cov}(\chi_{A_1}, \chi_{A_2}) = \mathbb{E}[\chi_{A_1}] - \mathbb{E}[\chi_{A_1}]\mathbb{E}[\chi_{A_2}] = p^3 - p^6. \tag{8}$$

We now put these results together to compute the variance of N_Δ. First we count the number of node triples A_1 and A_2 such that $\Delta(A_1) \cap \Delta(A_2)$ consists of a single edge. For this to happen $A_1 \cap A_2$ should consist of just 4 nodes. This can be done in $\binom{n}{4}$ ways, and these 4 nodes can be split into triples A_1 and A_2 in 4 ways. Any such split uniquely defines the common edge. Being sloppy about constants as usual, the number of ways this can happen is $c_1 n^4$. Likewise, the number of ways we can choose node triples A_1 and A_2 that coincide exactly is simply the number of ways we can choose A_1, which is $\binom{n}{3} = c_2 n^3$. Combining these counts with the covariance estimates in (7) and (8), and substituting in (6), we get

$$\mathrm{Var}(N_\Delta) = c_1 n^4 (p^5 - p^6) + c_2 n^3 (p^3 - p^6) = c_1 n^4 p^5 (1-p) + c_2 n^3 p^3 (1 - p^3).$$

Substituting $p = n^{-\alpha}$, and recalling that $\mathbb{E}[N_\Delta] = \binom{n}{3} p^3$, we see that

$$\frac{\mathrm{Var}(N_\Delta)}{(\mathbb{E}N_\Delta)^2} = \tilde{c}_1 n^{-2+\alpha}(1-p) + \tilde{c}_2 n^{-3(1-\alpha)}(1-p^3).$$

Now p is going to zero, so $1-p$ and $1-p^3$ are close to 1. Also, $\alpha < 1$ by assumption, so the exponents on n in the terms on the RHS are both strictly negative, so the RHS is tending to zero as n tends to infinity. Hence, we have shown that $\mathrm{Var}(N_\Delta)/(\mathbb{E}N_\Delta)^2$ tends to zero, as we set out to do. By Chebyshev's inequality, it follows that the probability of $G(n,p)$ containing no triangles tends to zero, and hence that it contains at least one triangle with high probability.

We are now ready to move on to the general case. But we need to restrict the class of graphs we consider in order to rule out ones like the counterexample described above. Given a graph H, denote by v_H and e_H the number of nodes and edges it has. Recall that a subgraph of H is a graph made up of a subset of the vertices and edges of H. For example, one of the subgraphs of a triangle is a V made up of all 3 nodes and any two edges; another is two

nodes and the edge between them; a third is all 3 nodes but just one edge; and there are many others. For a subgraph H' of H, we define $v_{H'}$ and $e_{H'}$ analogously. We define the density of a graph H as the ratio e_H/v_H.

Definition: A graph H is said to be *balanced* if it is at least as dense as every one of its subgraphs, i.e.,

$$\frac{e_H}{v_H} \geq \frac{e_{H'}}{v_{H'}} \text{ for all subgraphs } H' \text{ of } H.$$

We will state and prove a result that applies to balanced graphs, and then explain how to use it for unbalanced graphs, without giving a proof.

Theorem 1 *Let H be a balanced graph with v_H nodes and e_H edges. Consider a sequence of Erdős-Rényi random graphs $G(n, n^{-alpha})$, indexed by n, the number of nodes, and with fixed parameter α. Then,*

$$\mathbb{P}(G(n, n^{-\alpha}) \textit{ contains a copy of } H) \to \begin{cases} 0, & \textit{if } \alpha > v_H/e_H, \\ 1, & \textit{if } \alpha < v_H/e_H. \end{cases}$$

Proof. We shall estimate the mean and variance of N_H, the number of copies of H in $G(n,p)$, and then use Markov's and Chebyshev's inequalities as we did for the triangle. First, the mean is easy. Every subset of nodes A consisting of v_H nodes can support some constant number of copies of H. Let us denote this constant by c_H. (The constant is $v_H!$ divided by the size of the automorphism group of H.) Now, there are $\binom{n}{v_H}$ ways of choosing v_H nodes, and the probability that a particular copy of H appears is p^{e_H}, since H has e_H edges, each of which is independently present with probability p. Hence, we have

$$\mathbb{E}[N_H] = \binom{n}{v_H} c_H p^{e_H} = c'_H n^{v_H} p^{e_H}. \tag{9}$$

Next, we compute the variance of N_H. For each possible copy H_i that could appear in $G(n,p)$, let χ_{H_i} denote the indicator that all edges of H_i are present in $G(n,p)$. Clearly $\mathbb{E}[\chi_{H_i}] = p^{e_H}$, since H_i is present only if e_H edges are present. For two copies, H_i and H_j, we need to compute the covariance of the indicators of H_i and H_j. If the copies have no edges in common, then the covariance is zero. Otherwise, it depends on how many edges they have

in common. Suppose $H_i \cap H_j = H'$, where H' is necessarily a subgraph of H (possibly empty, and possibly all of H). Then,

$$\begin{aligned} \mathrm{Cov}(\chi_{H_i}, \chi_{H_j}) &= \mathbb{E}[\chi_{H_i}\chi_{H_j}] - \mathbb{E}[\chi_{H_i}]\mathbb{E}[\chi_{H_j}] \\ &= p^{2e_H - e_{H'}} - (p^{e_H})^2 = p^{2e_H - e_{H'}}(1 - p^{e_{H'}}). \end{aligned} \quad (10)$$

Next, we need to count the number of ways in which subgraphs H_1 and H_2 overlapping in H' could appear. The number of nodes required for this pattern is $v_{H_1} + v_{H_2} - v_{H'} = 2v_H - v_{H'}$, where v_H is the number of nodes in H (and also in H_1 and H_2, which are copies of H), and $v_{H'}$ the number of nodes in H'. (These are common nodes in H_1 and H_2 and have been counted twice in summing v_{H_1} and v_{H_2}, so we need to subtract this number from the sum.) The number of ways we can choose this many nodes is $\binom{n}{2v_H - v_{H'}}$. Now, having chosen these nodes, there are a number of possible ways that a pattern consisting of two copies of H intersecting in a copy of H' could appear on these nodes. Let us denote the number of ways by $c_{H,H'}$. While it is hard to calculate this number, it is clearly just a constant that does not depend on n or p. So, in our usual sloppy way, we won't bother calculating it. Putting together the expression for the covariance of indicators in (10) with our estimate for the number of ways such a term could occur, we obtain the following for the variance of N_H:

$$\mathrm{Var}(N_H) = \sum_{H' \subseteq H} \tilde{c}_{H,H'} n^{2v_H - v_{H'}} p^{2e_H - e_{H'}},$$

where the sum is taken over all subgraphs H' of H. This equation has been obtained by thinking of N_H as a sum of indicator random variables for the occurence of copies of H at different locations in $G(n,p)$, and then writing its variance as the sum of covariances for all pairs of indicators. The idea is exactly the same as for triangles, but we skipped writing this intermediate step explicitly.

Now, combining the above expression for the variance of N_H with the expression in (9) for its mean, we get

$$\frac{\mathrm{Var}(N_H)}{(\mathbb{E}N_H)^2} = \sum_{H' \subseteq H} \hat{c}_{H',H} n^{-v_{H'}} p^{-e_{H'}}, \quad (11)$$

where $\hat{c}_{H,H'}$ is some constant that does not depend on n or p.

We now take $p = n^{-\alpha}$. Suppose first that $\alpha > v_H/e_H$. Then, by (9), $\mathbb{E}[N_H] = c'_H n^{v_H - \alpha e_H}$ tends to zero as n tends to infinity, because the exponent on n is strictly negative. Hence, by Markov's inequality, $\mathbb{P}(N_H \geq 1)$ tends to zero, which yields the first claim of the theorem.

Suppose next that $\alpha < v_H/e_H$. By the assumption that H is balanced, we have $e_{H'}/v_{H'} \leq e_H/v_H$ for all subgraphs H' of H; in words, H is at least as dense (where density is defined as the ratio of edges to vertices) as any of its subgraphs. Hence, we also have $\alpha < v_{H'}/e_{H'}$ for any subgraph H' of H. Hence, it follows from (11) that

$$\frac{\mathrm{Var}(N_H)}{(\mathbb{E}N_H)^2} = \sum_{H' \subseteq H} \hat{c}_{H',H} n^{-(v_{H'} - \alpha e_{H'})}$$

tends to zero as n tends to infinity, as the exponent on n is strictly negative. Consequently, by Chebyshev's inequality, $\mathbb{P}(N_H = 0)$ tends to zero, i.e., $\mathbb{P}(N_H \geq 1)$ tends to 1. This completes the proof of the theorem. □

Remark. If H is not a balanced graph, then the value of α at which H appears in $G(n, n^{-\alpha})$ is determined by the densest subgraphs of H. If H' is one such densest subgraph (H could have more than one), then take $\alpha_c = v_{H'}/e_{H'}$. Even if the densest subgraph is not unique, the value of α_c is well defined because, by definition, all densest subgraphs have the same value for this ratio. The theorem applies to H', and says that the probability of H' appearing in $G(n, n^{-\alpha})$ is close to zero for $\alpha > \alpha_c$ and close to 1 for $\alpha < \alpha_c$. It turns out that the same is true for H. The appearance of the densest subgraphs is the 'bottleneck' for the appearance of H. As soon as $\alpha < \alpha_c$, copies of H' appear, and so do copies of H (possibly many more than of H'). On the other hand, if $\alpha > \alpha_c$, then H' doesn't appear in the random graph, and so H can't possibly appear.

2 The giant component

We have so far looked at "local properties" of Erdős-Rényi random graphs, by which we mean properties determined by a small number of vertices (a constant number, whereas the total number of vertices, n, tends to infinity). Now, we turn our attention to "global properties". Fix n large and a consider a sequence of Erdős-Rényi random graphs $G(n, p)$ indexed by p. As p increases from 0 to 1, the random graph evolves from the empty graph, consisting of isolated vertices, to the complete graph, in which all edges are present.

In previous sections, we looked at when certain subgraphs, such as triangles, appear in the random graph. Now, we want to ask how the size of the

largest connected component evolves with p. In the next section, we will ask when the whole graph becomes connected (i.e., there is a single connected component consisting of all n vertices).

It would be natural to expect that the size of the largest connected component increases smoothly from 1 to n as p increases from 0 to 1. However, this is not what happens. Instead, most components remain small, consisting only of a constant number of vertices (not growing with n) if p is smaller than $1/n$. But there is an abrupt change at $1/n$, when a component of size proportional to n (i.e., consisting of a fraction of all vertices) suddenly emerges. We now give a more careful statement of this result.

Theorem 2 *Consider a sequence of Erdős-Rényi random graphs $G(n, p_n)$, with $p_n = \lambda/n$. Let C_n denote the size of the largest connected component in $G(n, p_n)$. Then, the following holds with high probability (whp):*

- *If $\lambda \in [0, 1)$, then $C_n = O(\log n)$.*
- *If $\lambda > 1$, then $C_n \sim \rho_\lambda n$, where ρ_λ is the unique solution in $(0, 1)$ of the equation $e^{-\lambda x} = 1 - x$.*

Recall that the notation $a_n \sim b_n$ denotes that the ratio a_n/b_n tends to 1 as n tends to infinity. We won't give a full proof of this theorem, but will provide the intuition behind it, which comes from branching processes.

2.1 Review of branching processes

Branching processes are a model of stochastic population growth. Time is discrete, and time steps are called generations. We denote by Z_n the population size in the n^{th} generation. Each individual in the population in generation n has a random number of offspring and dies at the end of this generation. We can give a precise probabilistic description of this model as follows.

Let $\xi_{i,j}$, $i, j \in \mathbb{N}$ be iid random variables, taking values in $\mathbb{Z}_+ = \{0, 1, 2, \ldots\}$. Here, $\xi_{i,j}$ denotes the random number of offspring that the j^{th} individual in the i^{th} generation has (or would have, if this individual existed). Thus, we have

$$Z_{n+1} = \xi_{n,1} + \xi_{n,2} + \ldots + \xi_{n,Z_n}, \tag{12}$$

with the usual convention that an empty sum is equal to zero. The discrete time stochastic process Z_n, $n \in \mathbb{N}$ is called a Galton-Watson branching process. It is easy to see that Z_n is a (time homogeneous) Markov chain on the state space $\mathbb{Z}_+ = \{0, 1, 2, \ldots\}$.

We would like to know how Z_n behaves as n tends to infinity, and in particular whether or not it is non-zero, i.e., whether the population has become extinct by the n^{th} generation or not. The answer depends on the mean of the offspring distribution, which we denote by μ, i.e., $\mu = \sum_{k=0}^{\infty} k\mathbb{P}(\xi_{1,1} = k)$.

Lemma 1 *Consider a branching process $Z_n, n \in \mathbb{N}$, with some fixed initial condition $Z_1 = m$. Suppose that the offspring distribution has mean $\mu < 1$. Then, with probability one, the branching process eventually becomes extinct; in other words, Z_n tends to zero almost surely (*a.s.*) as n tends to infinity.*

Proof. We first compute the expectation of Z_n using (12). Using the fact that the $\xi_{i,j}$ are iid, and hence that $\xi_{n,j}, j \in \mathbb{N}$ are independent of Z_n, we have

$$\mathbb{E}[Z_{n+1}|Z_n] = \sum_{i=1}^{Z_n} \mathbb{E}[\xi_{n,i}] = \mu Z_n.$$

We have used the linearity of expectation to obtain the first equality, and the independence of $\xi_{n,i}$ from Z_n to get the second. Taking expectations again, we have

$$\mathbb{E}[Z_{n+1}] = \mathbb{E}(\mathbb{E}[Z_{n+1}|Z_n]) = \mu\mathbb{E}[Z_n].$$

Applying this equality recursively, we get $\mathbb{E}[Z_n] = m\mu^{n-1}$. Now, it follows from Markov's inequality that

$$\mathbb{P}(Z_n \neq 0) = \mathbb{P}(Z_n \geq 1) \leq \frac{\mathbb{E}[Z_n]}{1} = m\mu^{n-1}.$$

Since $\mu < 1$, it is clear that $\mathbb{P}(Z_n \neq 0)$ tends to zero as n tends to infinity.

Now, the events $\{Z_n \neq 0\}$ are decreasing, as their complements are increasing: $Z_n = 0$ implies that $Z_{n+1} = 0$, or equivalently, $\{Z_n = 0\} \subseteq \{Z_{n+1} = 0\}$. The event that the branching process eventually becomes extinct is the event $\{\exists n : Z_n = 0\} = \cup_{n=1}^{\infty}\{Z_n = 0\}$. Consequently, the probability of non-extinction is bounded by

$$\mathbb{P}\Big(\cap_{n=0}^{\infty}\{Z_n \geq 1\}\Big) = \lim_{n\to\infty} \mathbb{P}(Z_n \geq 1) \leq \lim_{n\to\infty} m\mu^{n-1} = 0.$$

This completes the proof of the lemma. □

Suppose $Z_1 = 1$, i.e., we start with a single individual. Denote the generating function of the offspring distribution by

$$G(u) = \mathbb{E}[u^{\xi_{1,1}}] = \sum_{k=0}^{\infty} u^k \mathbb{P}(\xi_{1,1} = k). \tag{13}$$

Clearly, this is also the generating function of Z_2, the population size in the second generation, as

$$Z_2 = \sum_{j=1}^{Z_1} \xi_{1,j} = \xi_{1,1}.$$

Using the fact that the $\xi_{i,j}$ are iid, and hence that $\xi_{n,j}, j \in \mathbb{N}$ are independent of Z_n, we can recursively compute the generating function of Z_{n+1} as follows:

$$\begin{aligned} G_{n+1}(u) &:= \mathbb{E}[u^{Z_{n+1}}] = \mathbb{E}(\mathbb{E}[u^{Z_{n+1}}|Z_n]) \\ &= \Big(\sum_{k=0}^{\infty} \mathbb{P}(Z_n = k)\mathbb{E}\Big[u^{\sum_{j=1}^{Z_n}\xi_{n,j}} \;\Big|\; Z_n = k\Big]\Big) \\ &= \Big(\sum_{k=0}^{\infty} \mathbb{P}(Z_n = k)G(u)^k\Big) = G_n(G(u)). \end{aligned}$$

In other words, the generating function of Z_n is $G_n(\cdot) = G^{(n)}(\cdot) := G \circ G \circ \cdots \circ G(\cdot)$, the n-fold composition of the generating function G of the offspring distribution. We shall use this fact to prove the next result about when branching processes survive forever with positive probability.

Theorem 3 *Consider a branching process $Z_n, n \in \mathbb{N}$, with initial condition $Z_1 = 1$, and let $G(\cdot)$ denote the generating function of the offspring distribution. Then, the branching process eventually becomes extinct with probability p_e which is the smallest root in $[0, 1]$ of the equation $G(x) = x$; it survives forever with the residual probability, $1-p_e$. Moreover, $p_e = 1$ if $\mu = \mathbb{E}[\xi] < 1$ and $p_e < 1$ if $\mu > 1$.*

*P*roof. Let us begin by recalling some properties of generating functions. By definition,

$$G(x) = \sum_{n=-\infty}^{\infty} \mathbb{P}(\xi = n)x^n.$$

Now, $x \mapsto x^n$ is a convex function on $\mathbb{R}_+ = [0, \infty)$ for all $n \in \mathbb{Z}$. Since $G(\cdot)$ is a linear combination of convex functions with non-negative coefficients, it is also a convex function. (We have not used the fact that $\mathbb{P}(\xi = n) = 0$ for $n < 0$. Generating functions of all discrete random variables are convex, whether or not the random variables are non-negative.) It is easy to see that $G(1) = 1$ and, using the non-negativity of ξ, that $G(0) = \mathbb{P}(\xi = 0)$. In particular, $G(0) \geq 0$.

We shall also use the fact that $G(\cdot)$ is continuous on $[0, 1]$ (which follows from its convexity), and that $\mathbb{E}[\xi] = G'(1)$ if $G(\cdot)$ has a finite left derivative at 1, and $\mathbb{E}[\xi] = +\infty$ otherwise. Now, it is not hard to see that the equation $G(x) = x$ has a unique solution in $[0, 1)$ if $G'(1) > 1$, and no solution in $[0, 1)$ if $G'(1) < 1$. (It might be useful to draw a picture.) Thus, the smallest solution of $G(x) = x$ on $[0, 1]$, which we denote x^*, satisfies $x^* < 1$ in the former case, and $x^* = 1$ in the latter. Also $G(x) > x$ on $(0, x^*)$ and $G(x) < x$ on $(x^*, 1)$; if $x^* = 1$, the latter interval is empty, and the claim is vacuous.

Suppose first that $G'(1) < 1$, and consequently that $\mu := \mathbb{E}[\xi] < 1$. Then, $x^* = 1$. Moreover, we showed in Lemma 1 that extinction is certain if $\mu < 1$. Thus, $p_e = 1$, and the theorem is proved in this case.

Suppose next that $G'(1) > 1$. We saw above that the generating function of Z_n is given by $G^{(n)}$, the n-fold composition of G with itself. Suppose $x \in [0, x^*)$. Then, $G(x) > x$. Also, $G(x) < x^*$ because G is a monotone increasing function, as ξ is non-negative. Hence, $G(x) \in (x, x^*)$. Repeating this argument, the sequence $G^{(n)}(x)$, $n = 1, 2, 3, \ldots$ is a monotone increasing sequence, bounded above by x^*. Hence, it converges to a limit, and we claim that the limit must be x^*. If it were some $y < x^*$, then we would have that $G(y) = y$ (by continuinty of G), which violates the definition of x^* as the smallest root of $G(x) = x$ on $[0, 1]$. In particular, $G^{(n)}(0)$ tends to x^*. But $G^{(n)} = \mathbb{P}(Z_n = 0)$. Thus, we have shown that $\mathbb{P}(Z_n = 0)$ tends to x^*. Consequently, the extinction probability is given by

$$p_e = \mathbb{P}\Big(\cup_{n=0}^{\infty}\{Z_n = 0\}\Big) = \lim_{n\to\infty} \mathbb{P}(Z_n = 0) = x^*.$$

This completes the proof of the theorem. □

2.2 Emergence of the giant component

We now return to the study of the component structure of the Erdős-Rényi random graph, $G(n,p)$. Fix a vertex v and consider a process that explores expanding neighbourhoods of v. Let S_1 denote the set of neighbours of v, S_2 the set of nodes at distance 2 from v, and so on. Here, distance refers to graph distance: the smallest number of edges on any path between two nodes. Let $Z_1, Z_2, \ldots$ denote the cardinalities of the sets $S_1, S_2, \ldots$. As the edge between v and any other node w is present with probability p, independent of all other edges, it is clear that $Z_1 \sim \text{Bin}(n-1,p)$. The notation means that the random variable Z_1 has a Binomial distribution with parameters $n-1$ and p. Recall the scaling regime we are considering, in which n tends to infinity and p tends to zero as $p = \lambda/n$ for a fixed constant $\lambda > 0$ that doesn't depend on n. In this scaling regime, $\text{Bin}(n-1,p) \approx \text{Pois}(\lambda)$, i.e., Z_1 is approximately Poisson distributed with parameter (mean) λ.

Next, for each node u in S_1, the number of neighbours of u excluding v is binomially distributed with parameters $n-2$ and p, which is also approximately $\text{Pois}(\lambda)$. Now, not all of these nodes may lie at distance 2 from v as some of them may already belong to the set S_1. However, as there are $n-2$ nodes for u to choose from as neighbours, and only about λ of them belong to S_1, the chances of u choosing one of them are negligible. In addition, if u and w are two different nodes in S_1, the chance that both of them will choose the same node as their neighbour is also negligible. Thus, each node in S_1 contributes approximately $\text{Pois}(\lambda)$ nodes to S_2. Similar reasoning applies to S_3, S_4 and so on, until these sets grow big enough that they contain some fraction of the total number of nodes, n.

What the above description tells us is that the sequence of random variables $Z_1, Z_2, \ldots$ behaves approximately like the population size in successive generations of a Galton-Watson branching process with $\text{Poisson}(\lambda)$ offspring distribution. (Note: There is a slight change of notation from the previous subsection. Let $S_0 = \{v\}$ and $Z_0 = 1$. Then, $Z_0, Z_1, Z_2 \ldots$ here are the same as $Z_1, Z_2, Z_3, \ldots$ in the previous subsection.) We can be more precise and provide bounds: the random variables $Z_0, Z_1, Z_2, \ldots$ denoting neighbourhood sizes are stochastically dominated by the population sizes in a branching process with $\text{Bin}(n,p)$ offspring distribution. Morever, until the first generation that $Z_0 + Z_1 + \ldots + Z_k \geq \epsilon n$, where $\epsilon > 0$ is a given constant, the random variables Z_i stochastically dominate the population size

in a branching process with $\text{Bin}(n, (1-\epsilon)p$ offspring distribution. These observations can be used to provide a rigorous proof of Theorem 2. However, we shall confine ourselves to just providing the intuition behind it.

Consider again the exploration process starting from a given vertex v, finding its successive neighbourhoods $S_1(v), S_2(v), \ldots$, where we have now made the dependence of these neighbourhoods on v explicit in the notation. The process ends if, for some k, $S_{k+1}(v) = \emptyset$. At this point, we have found all the nodes that can be reached from v, which we call the cluster (or connected component) containing v, and denote $C(v)$. Thus,

$$C(v) = \bigcup_{i=0}^{k} S_i(v) \text{ and } |C(v)| = \sum_{i=0}^{k} Z_i(v).$$

Now, we use the intuition that $Z_i(v)$ is approximately the population size in the i^{th} generation of a branching process with Poisson(λ) offspring distribution. If $\lambda < 1$, then this branching process is guaranteed to go extinct, and $|C(v)|$ has roughly the distribution of the total population size in the branching process until extinction. This is some random variable parametrised by λ, and doesn't depend on n or p directly. Thus, the cluster containing v is O(1) in size, in probability: given any $\epsilon > 0$, we can find a constant M_ϵ large enough (depending only on ϵ and not on n) such that $\mathbb{P}(|C(v)| > M_\epsilon) < \epsilon$, uniformly in n.

We can repeat this process to find all the connected components of the random graph $G(n, p)$. Start from some vertex v_1 and identify its cluster $C(v_1)$. If this cluster doesn't contain all nodes, then choose a vertex $v_2 \notin C(v_1)$, and identify its cluster, $C(v_2)$. Repeat the process until $\cup_{i=1}^{k} C(v_i) = V$. Our reasoning above applies to each of these clusters. In fact, as the exploration proceeds, the number of nodes left keeps getting smaller, and so do the binomial random variables denoting the neighbourhood sizes. Consequently, a Poisson(λ) random variable continues to stochastically dominate them, and cluster sizes remain $O(1)$ in probability. The upper bound of $O(\log n)$ in the statement of Theorem 2 comes from taking the maximum of the sizes of the clusters $C(v_1)$, $C(v_2)$ and so on. Justifying it would require us to derive bounds on the population size of a subcritical branching process with Poisson offspring distribution, showing that the probability of large population sizes decays exponentially. This can be done using the generating function approach developed above, but we will not do it here.

Next, let us turn to the case $\lambda > 1$. Theorem 3 tells us that a branching process with Poisson(λ) offspring distribution becomes extinct with probability

p_e which is the smallest root in $[0, 1]$ of the equation

$$e^{-\lambda+\lambda x} = x.$$

It survives forever with the residual probability $1 - p_e$, which we shall denote by ρ_λ to make its dependence on λ explicit. Now, if we consider the neighbourhood exploration process started from a vertex v, the comparison with the branching process tells us that the exploration terminates soon with probability $p_e = 1 - \rho_\lambda$, and yields a cluster $C(v)$ which is $O(1)$ in size, as before. As this reasoning applies to any starting vertex v, it implies that the expected number of vertices in small components is $(1 - \rho_\lambda)n$.

But, with probability ρ_λ, the branching process survives forever, and the population size tends to infinity. Obviously, the cluster sizes are bounded above by n and cannot grow to infinity for any fixed n. What happens instead is that the cluster gets so large that its growth can no longer be well approximated by a branching process. With a bit of care, and the bounding techniques mentioned earlier, this argument can be used to show that the big clusters grow to size at least ϵn for some sufficiently small $\epsilon > 0$.

So, what we have obtained so far (somewhat loosely) is that $(1 - \rho_\lambda)n$ vertices, on average, belong to small clusters, while $\rho_\lambda n$ belong to one or more large clusters. Theorem 2 makes the stronger claim that, in fact, these $\rho_\lambda n$ vertices all belong to a single giant cluster. To see this, we argue as follows. Consider the cluster growth (or neighbourhood exploration) processes started from two different vertices v_1 and v_2, and continued until they have grown to ϵn nodes (assuming that v_1 and v_2 belong to big clusters, and hence these processes don't become extinct before that). If the two clusters have already intersected by this time, then v_1 and v_2 belong to the same cluster, and we are done. Otherwise, the important point to note is that there some $\epsilon' n$ nodes which lie on the boundary of each cluster, and whose neighbourhoods are still unexplored. Here, ϵ' is some constant in $(0, \epsilon)$. The reason this is true is that cluster sizes (or branching processes in the supercritical case) grow geometrically, and so the last / most recent generation contains a non-vanishing fraction of the total population that ever lived. Now, our claim is that, with high probability, there must be at least one edge between these boundary nodes. Indeed, as each boundary is of size $\epsilon' n$, there are $(\epsilon')^2 n^2$ potential edges, each of which is present with probability $p = \lambda/n$, independent of the others. Hence, the probability that all these edges are absent is given by

$$\Big(1 - \frac{\lambda}{n}\Big)^{(\epsilon')^2 n^2} \approx \exp\big(-\lambda(\epsilon')^2 n\big).$$

This probability tends to zero very quickly; even when we take a union bound over all pairs of possible starting vertices v_1 and v_2, the probability of finding two large clusters that don't intersect is vanishing. This completes the (hand-waving) proof of Theorem 3.

3 Connectivity

In this section, we want to answer the following question: how large does p have to be in order for the random graph $G(n,p)$ to be connected with high probability, i.e., to consist of a single connected component containing all the vertices? We saw in the last section that the random graph $G(n,\lambda/n)$ possesses a large connected component if $\lambda > 1$, and that the number of nodes in this giant component is approximately $\rho_\lambda n$, where ρ_λ solves the equation $e^{-\lambda x} = 1 - x$. It is not hard to see that ρ_λ is strictly smaller than 1 for any fixed $\lambda > 0$, however large. What this tells us is that the scaling regime in which $p = O(1/n)$ is not sufficient to yield full connectivity. We need to look at a regime in which p doesn't decrease to zero quite so fast as n tends to infinity. It turns out that the correct scaling regime to consider is p of order $(\log n)/n$. We will first state the result and then provide some intuition about it, and the outline of a proof.

Theorem 4 *Consider a sequence of Erdős-Rényi random graphs $G(n,p)$ indexed by n, with $p = (c\log n)/n$. We then have the following:*

$$\mathbb{P}(G(n,p) \text{ is connected}) \to \begin{cases} 1, & \text{if } c > 1, \\ 0, & \text{if } c < 1. \end{cases}$$

It turns out that connectivity also exhibits a sharp threshold, like the other properties we studied, namely the emergence of small subgraphs or motifs, and the emergence of the giant component. In other words, there is not a gradual increase in the probability that $G(n,p)$ is connected as p increases. Instead, for values of p slightly smaller than the threshold, the random graph is disconnected whp, whereas for p slightly larger than the threshold, it is connected whp.

Before going on to a proof of the theorem, let us describe the intuitive picture of what happens near the connectivity threshold. It turns out for p close to, but slightly smaller than, $(\log n)/n$, almost all the nodes belong to a

single connected component, but there are a small number of isolated nodes - nodes that have no edges to any other node. The number of isolated nodes is of order 1, whereas the connected component contains all n nodes except for this constant number. As p increases further, edges appear between these isolated nodes and the giant component (this is far more likely than that one appears between two isolated nodes) until, eventually, no more isolated nodes are left. This picture suggests the following approach to tackling the proof. Let us ask how likely it is that the random graph $G(n,p)$ contains isolated nodes, and how large p needs to be to ensure that with high probability there are no isolated nodes.

Lemma 2 *Let N denote the random number of isolated nodes in the random graph $G(n,p)$. Suppose $p = (c\log n)/n$. Then, $N = 0$ whp if $c > 1$, whereas $N \geq 1$ whp if $c < 1$.*

Proof. The proof uses the first and second moment methods, which we learnt when studying motifs. Letting χ_v denote the indicator that node v is isolated, we can express the random variable N as

$$N = \sum_{v \in V} \chi_v. \tag{14}$$

For a node v to be isolated, all $n-1$ possible edges from that node to the remaining nodes must be absent; this event has probability $(1-p)^{n-1}$. Consequently, using the linearity of expectation, we obtain

$$\mathbb{E}[N] = \sum_{v \in V} \mathbb{E}[\chi_v] = n\Big(1 - \frac{c\log n}{n}\Big)^{n-1} \sim ne^{-c\log n} = n^{1-c}, \tag{15}$$

where $f(n) \sim g(n)$ denotes that $f(n)/g(n)$ tends to 1 as n tends to infinity. Hence, if $c > 1$, then $\mathbb{E}[N]$ tends to zero as n tends to infinity, and so does $\mathbb{P}(N \geq 1)$ by Markov's inequality. It follows that $N = 0$ whp if $c > 1$. This proves the first claim of the lemma.

It also follows from equation (15) that the expected number of isolated nodes tends to infinity as n tends to infinity if $c < 1$. However, this is not enough to guarantee that there is at least 1 with high probability. To show that, we need to compute the variance of the number of isolated nodes and use the Second Moment Method. Now, the probability that two distinct nodes u and v are isolated is given by the probability that the edge between them

is absent and also that all $n-2$ possible edges from each of them to the remaining nodes is absent. In total, it requires $2n-3$ edges to be absent, which has probability p^{2n-3}. Hence, $\mathbb{E}[\chi_u\chi_v] = (1-p)^{2n-3}$, and

$$\mathrm{Cov}(\chi_u, \chi_v) = \mathbb{E}[\chi_u\chi_v] - \mathbb{E}[\chi_u]\mathbb{E}[\chi_v] = (1-p)^{2n-3} - (1-p)^{2n-2} = p(1-p)^{2n-3}.$$

On the other hand, if $u = v$, then $\chi_u\chi_v = \chi_u^2 = \chi_u$, and so

$$\mathrm{Var}(\chi_u) = \mathbb{E}[\chi_u^2] - (\mathbb{E}[\chi_u])^2 = (1-p)^{n-1} - (1-p)^{2(n-1)}.$$

Since $N = \sum_{v\in V} \chi_v$, it follows that

$$\begin{aligned} \mathrm{Var}(N) &= \sum_{u,v\in V} \mathrm{Cov}(\chi_u, \chi_v) \\ &= n(1-p)^{n-1}\big(1-(1-p)^{n-1}\big) + n(n-1)p(1-p)^{2n-3}. \end{aligned}$$

This is because there are n terms in the sum with $u = v$ and $n(n-1)$ terms with u and v different. Now, substituting $p = c(\log n)/n$, we get $(1-p)^n \sim e^{-c\log n} = n^{-c}$, and so

$$\mathrm{Var}(N) \sim n^{1-c} + nc\log n n^{-2c} \sim n^{1-c}. \tag{16}$$

Now, using equations (15) and (16), and applying Chebyshev's inequality, we get

$$\mathbb{P}(N=0) \le \mathbb{P}(|N - \mathbb{E}[N]| \ge \mathbb{E}[N]) \le \frac{\mathrm{Var}(N)}{(\mathbb{E}N)^2} \sim n^{c-1}. \tag{17}$$

Hence, if $c < 1$, then $\mathbb{P}(N=0)$ tends to 0, i.e., $N \ge 1$ whp. This proves the second claim of the lemma. □

The lemma suffices to prove one half of the theorem. If $c < 1$, then with high probability there is at least one isolated node. But if there is an isolated node, then the graph can't be connected. So, with high probability, $G(n, c\log n/n)$ is not connected if $c < 1$.

In order to prove the other direction, we need to work harder. We need to show that not only are there no isolated nodes, but that there no connected components (clusters) of any size up to $n/2$ that are disconnected from the rest of the graph. (Why is it enough to show this for clusters of size up to $n/2$ and not $n-1$?)

Fix $k \ge 2$. We will now obtain an upper bound on the probability that there is a cluster of size k disconnected from the rest of the graph. First, choose

k vertices from V. What is the probability that there are no edges between this subset and the remaining vertices? As there $k(n-k)$ potential edges, this probability is just

$$(1-p)^{k(n-k} = \Big(1 - \frac{c\log n}{n}\Big)^{k(n-k)} \sim \exp\Big(-\frac{ck(n-k)\log n}{n}\Big).$$

Next, what is the probability that these k nodes form a connected component? If they do, then the connected component must contain a spanning tree on k nodes, i.e., at least $k-1$ edges must be present. Hence, any given spanning tree has probability p^{k-1}. But we need to consider all possible spanning trees, which requires knowing how many of them there are. Luckily, there is a famous formula in combinatorics, known as Cayley's formula, which enumerates the number of trees on k labelled vertices. The formula says that there are exactly k^{k-2} such trees. (Check this for $k = 3$ and 4.) As each tree has probability p^{k-1} of being present, the union bound yields an upper bound on the probability that the given k-vertex subset is connected; the probability is no more than $k^{k-2}p^{k-1}$. As there are $\binom{n}{k}$ ways of choosing the k-vertex subset, we conclude that the probability of there being a cluster of size k disconnected from the rest of the graph is bounded above by

$$\binom{n}{k} k^{k-2}p^{k-1}(1-p)^{k(n-k)} \sim \binom{n}{k}\ k^{k-2}p^{k-1}\exp\Big(-\frac{ck(n-k)\log n}{n}\Big).$$

If we can show that the sum of the above expression over k ranging between 2 and $n/2$ tends to zero whenever $c > 1$, then we have completed the proof of the theorem. (The case $k = 1$ has already been dealt with by Lemma 2.) This calculation is not particularly deep, but it is messy and involved, and we will skip it.