1.7 Threshold for Connectivity

In this section we will investigate the question: How large does \( \lambda \) have to be so that the probability \( \text{ER}(n, \lambda/n) \) is connected (i.e., ALL vertices in ONE component) tends to 1. Half of the answer is easy. Let \( d_x \) be the degree of \( x \).

\[
P(d_x = 0) = \left(1 - \frac{\lambda}{n}\right)^n
\]

Using the series expansion of \( \log(1 - x) \) it is easy to see that if \( \lambda = o(n^{1/2}) \) then

\[
\left(1 - \frac{\lambda}{n}\right)^n e^\lambda \to 1
\]

(1.7.1)

Thus when \( \lambda = a \log n \) we have \( P(d_x = 0) \sim n^{-a} \). Thus if \( a < 1 \) the number of isolated vertices \( I_n = |\{x : d_x = 0\}| \) has

\[
EI_n = nP(d_x = 0) \sim n^{1-a} \to \infty
\]

To show that the actual value of \( I_n \) is close to the mean we note that if \( x \neq y \)

\[
P(d_x = 0, d_y = 0) = \left(1 - \frac{\lambda}{n}\right)^{2n-1} = \left(1 - \frac{\lambda}{n}\right)^{-1} P(d_x = 0)P(d_y = 0)
\]

so we have

\[
\text{var}(I_n) = nP(d_1 = 0)(1 - P(d_1 = 0)) + n(n-1) \left(1 - \left(\frac{\lambda}{n}\right)^{-1} - 1\right) P(d_1 = 0)P(d_2 = 0)
\]

When \( \lambda = a \log n \)

\[
\text{var}(I_n) \sim n^{1-a} + n^2 \left(\frac{\lambda}{n}\right) n^{-2a} \sim EI_n
\]

Using Chebyshev’s inequality it follows that if \( a < 1 \)

\[
P\left(|I_n - EI_n| > \omega(n)(EI_n)^{1/2}\right) \leq \frac{1}{\omega(n)^2}
\]

(1.7.2)

The last result shows that if \( \lambda = a \log n \) with \( a < 1 \) then with high probability there are about \( n^{1-a} \) isolated vertices, and hence the graph is not connected. Showing that the graph is connected is more complicated because we have to consider all possible ways in which the graph can fail to be connected. (1.5.1) and (1.7.1) tell us that the expected number of trees of size \( k \) is

\[
\sim n \frac{k^{k-2}}{k!} \lambda^{k-1} e^{-\lambda k}
\]
When \( k = 2 \) and \( \lambda = a \log n \) this is
\[
\frac{n}{2} (a \log n) n^{-2a}
\]
Thus if \( 1/2 < a < 1 \) there are isolated vertices, but no components of size 2. It is easy to generalize the last argument to conclude that when \( \lambda = a \log n \) and \( 1/(k+1) < a < 1/k \) there are trees of size \( k \) but not of size \( k+1 \). Bollobás (2001), see Section 7.1, uses this observation with the fact we know that the largest component is \( O(\log n) \) to sum the expected values to prove.

**Theorem 1.7.1.** Consider \( G = ER(n, \lambda/n) \) with \( \lambda = a \log n \). The probability \( G \) is connected tends to 0 if \( a < 1 \) and to 1 if \( a > 1 \).

**Proof.** We will use the approach of Section 1.3 to show that the probability a vertex fails to connect to the giant component is \( o(1/n) \). Since we have a large \( \lambda \) we can use the lower bound process with \( \delta = 1/2 \). The constant \( \theta_{1/2} \) that appears in (1.3.7) is defined by
\[
\theta + (\lambda/2)(e^{-\theta} - 1) = 0
\]
This is hard to compute for fixed \( \lambda \), so instead we decide we want \( \theta = 1 \) and see that this means \( \lambda = 2e/(e-1) \). Using monotonicity we see that if \( \lambda \geq 2e/(e-1) \), (1.3.8) implies that for our comparison random walk
\[
P_{2\log n}(T_0 < \infty) \leq n^{-2}
\]
To reach size \( 2 \log n \) we use the large deviations bound in Lemma 1.3.3, with \( \delta = 0 \) and \( x = 1/2 \) to conclude that if \( S_K \) is a sum of \( K \) independent Binomial\((n, \lambda/n)\) random variables then \( \mu = K \lambda \) and
\[
P(S_K \leq \mu/2) \leq \exp(-\gamma(1/2)\mu) \tag{1.7.3}
\]
where \( \gamma(1/2) = (1/2)\log(1/2) + 1/2 = 1/2(1 - \log(2)) \geq .15 \). If \( \lambda = (1+\epsilon) \log n \) with \( \epsilon \geq 0 \) and \( K = 14 \) then
\[
P(S_{14} \leq 7 \log n) \leq n^{-2.1}
\]
The last calculation shows that if \( x \) has at least 14 neighbors, then with high probability in two steps it can reach at least \( 7 \log n \) vertices. The next step is to bound
\[
P(d_x \leq 13) = \sum_{k=0}^{13} \binom{n}{k} \left( \frac{\lambda}{n} \right)^k \left( 1 - \frac{\lambda}{n} \right)^{n-k}
\]
\[
\leq \sum_{k=0}^{13} \frac{\lambda^k}{k!} e^{-\lambda(n-k)/n} \leq 14 (a \log n)^{13} n^{-a} e^{(13a \log n)/n}
\]
which is \( o(n^{-1}) \) if \( a > 1 \).
1.7. THRESHOLD FOR CONNECTIVITY

To finish up now (and to prepare for the next proof), we apply the large deviations result lemma 1.3.3 to lower bounding random walk \( W_t \) twice to conclude that if 
\[-1 + (\alpha \log n) \geq (1/5) \log n \]
then there are positive constants \( \eta_i \) so that
\[
P(W(n^{1/2}) - W(0) \leq 0.1n^{1/2} \log n) \leq \exp(-\eta_1 n^{1/2} \log n)
\]
\[
P(W(n^{1/2}) + n^{1/2} - W(0) \geq 2an^{1/2} \log n) \leq \exp(-\eta_2 n^{1/2} \log n)
\]

Combining our results we see that with probability \( 1 - o(n^{-1}) \) the RAU process will not expose \( n/2 \) vertices and have at least \( 0.1n^{1/2} \log n \) active vertices at time \( n^{1/2} \). When this occurs for \( x \) and for \( y \) the probability that their clusters fail to intersect is at most
\[
\left( 1 - \frac{\log n}{n} \right)^{0.01n(\log n)^2} \leq e^{-\log n^3/100}
\]
and the proof is complete.

The next result is an easy extension of the previous argument and will allow us to get a sharper result about the transition to connectivity.

**Theorem 1.7.2.** Consider \( G = ER(n, \lambda/n) \) with \( \lambda = a \log n \). If \( a > 1/2 \) then with probability tending to 1 \( G \) consists only of a giant component and isolated vertices.

**Proof.** It follows from the previous argument that if \( n \) is large \( P_{2\log n}(T_0 < \infty) \leq n^{-2} \). Using (1.7.3) it is easy to see that if \( S_{28} \) is a sum of 28 independent Binomial(\( n, \lambda/n \)) and \( \lambda \geq (1/2) \log n \) then
\[
P(S_{28} \leq 7 \log n) \leq n^{-2.1}
\]

Consider now a branching process \( Z_k^x \) with offspring distribution Binomial(\( n, \lambda/n \)). If \( Z_1^x = 0 \) the cluster containing \( x \) is a singleton. We might be unlucky and have \( Z_1^x = 1 \) but in this case
\[
P(Z_1^x = 1, Z_2^x \leq 27) = n \frac{\lambda}{n} \left( 1 - \frac{\lambda}{n} \right)^{n-1} \sum_{k=0}^{27} \binom{n}{k} \left( \frac{\lambda}{n} \right)^k \left( 1 - \frac{\lambda}{n} \right)^{n-k}
\]
\[
\leq \sum_{k=0}^{27} \frac{\lambda^{k+1}}{k!} e^{-\lambda(2n-k-1)/n} \leq 28(a \log n)^{28} n^{-2a} e^{(28\alpha \log n)/n}
\]

which is \( o(n^{-1}) \) if \( a > 1/2 \). If we are lucky enough to find two neighbors on the first try then
\[
P(Z_2^x \leq 28 | Z_1^x \geq 2) \leq \frac{28}{k=0} \binom{2n}{k} \left( \frac{\lambda}{n} \right)^k \left( 1 - \frac{\lambda}{n} \right)^{2n-k}
\]
\[
\leq \sum_{k=0}^{28} \frac{(2\lambda)^k}{k!} e^{-2\lambda(n-k)/n} \leq 29(a \log n)^{28} n^{-2a} e^{(56\alpha \log n)/n}
\]
Here we have replaced 27 by 28 to take care of collisions. One collision has probability \( c/n \), but by a now familiar estimate, the probability of two collisions in one step of the branching process is \( O(1/n^2) \).

In the previous proof we only used \( a > 1/2 \) in the W-estimates and the final estimate so the proof is complete.

We are now ready to more precisely locate the connectivity threshold.

**Theorem 1.7.3.** Consider \( G = ER(n, \lambda/n) \) with \( \lambda = \log n + b + o(1) \). The number of isolated vertices \( I_n \) converges to a Poisson distribution with mean \( e^{-b} \) and hence the probability \( G \) is connected tends to \( \exp(-e^{-b}) \).

**Proof.** By the previous result \( G \) will be connected if and only if there are no isolated vertices. Using 1.7.1), the probability \( x \) is isolated is

\[
\left(1 - \frac{\lambda}{n}\right)^n \sim \exp(-\log n - b) \sim e^{-b}/n
\]

so \( EI_n \). The expected number of ordered \( k \)-tuples of isolated vertices is

\[
(n \cdot (n - 1) \cdots (n - k + 1)) \left(1 - \frac{\lambda}{n}\right)^{n+(n-1)+\cdots+(n-k+1)} \rightarrow e^{-bk}
\]

so the Poisson convergence follows from the method of moments. \( \square \)