Chapter 3

Inhomogeneous random graphs

In maximum generality an inhomogeneous random graph is one in which an edge between vertices $i$ and $j$ is present with probability $p(i, j)$. However, it is not possible to prove anything significant in this level of generality. Bollobás, Janson, and Riordan (2007), a reference we will refer to as BJR(2007), have found a framework that is very general and allows for a rich theory to be developed. In the special case that we will consider, we imagine the vertices embedded in $(0, 1]$ which is equipped with the Borel sets and Lebesgue measure, and there is a kernel $\kappa$ defined on $(0, 1] \times (0, 1]$ (with properties to be specified later) so that

$$p(i, j) = \kappa(i/n, j/n)/n$$

We will describe the set-up in Section 3.3 after we have introduced a number of examples in Section 3.2 to show that this framework is useful. In the next section we will introduce the situation in which there are only a finite number of different types of vertices. This is the simplest possible example but it is important, because results for this case are an important first step for proving things in general.

3.1 Finitely many types

For each $1 \leq k \leq m$, there $n_k$ individuals of type $k$, and there is an edge from an individual of type $j$ to one of type $k$ with probability $a(j, k)/n$ where $n = \sum_k n_k$. If we let $\ell(x)$ be the type of vertex $x$ then

$$p(x, y) = \frac{a(\ell(x), \ell(y))}{n}$$

It is natural to arrange the indices so that the first $n_1$ vertices are of type 1, the next $n_2$ are of type 2, etc and to assume that $n_k/n \to \pi_k$. Let $V_k$ be the set of vertices of type $k$.

When there is only one type and $a(1, 1) = \lambda$, we have a Erdős-Rényigraph. In the limit $n \to \infty$ vertices have degrees that are Poisson with mean $\lambda$. In Section 1.2 we compared the growth of clusters to a branching process in which individuals have a Poisson($\lambda$) number of offspring and concluded that
(i) there is a giant component (with size of order \( n \)) if and only if \( \lambda > 1 \).

(ii) if \( \lambda > 1 \) then as \( n \to \infty \) the fraction of vertices in the giant component converges to the survival probability of the branching process.

In the multitype case, if we let \( N_{j,k} \) be the number of neighbors of type \( k \) for a vertex of type \( j \) then (i) the \( N_{j,k} \) have a distribution that is

\[
\text{binomial}(n_k, a(j,k)/n) \approx \text{Poisson}(a(j,k)\pi_k)
\]

and (ii) \( N_{j,1} \ldots N_{j,m} \) are independent.

Our first goal is to generalize results for the Erdős-Rényi case and show that the phase transition can be understood by studying a multitype branching process. Let \( \mu(j,k) = EN_{j,k} \).

If there is \( \ell \) so that if \( \mu_{jk}^\ell > 0 \) for all \( j,k \) (in which case Athreya and Ney (1972) say the branching process is \textbf{positive regular}) then the \textbf{Perron-Frobenius theorem} implies

\[
\text{Theorem 3.1.1. As } t \to \infty \\
\mu_{j,k}^t \sim \alpha^t v_j u_k
\]

where \( \alpha \) is the largest eigenvector of \( \mu \), and \( u, v \) are the associated left \((1 \times m)\) and right \((m \times 1)\) eigenvectors normalized so that \( u \cdot v = 1 \).

The maximum eigenvalue is often denoted by \( \rho \) but here that letter is reserved for the survival probability of the branching process.

It may seem strange to have the right eigenvector on the left in the limit in (3.1.1), but this is dictated by the fact that \( vu \) is a \( m \times m \) matrix. As a further check, note that in the case \( \mu(j,k) \) is a Markov chain transition probability (i.e., the rows sum to 1), the maximum eigenvalue \( \alpha = 1 \), the left eigenvector \( u \) is the stationary distribution while the right eigenvector \( v \) is a vector of all 1’s, so \( v_i u_j \) is a matrix in which each row is the stationary distribution.

\[\text{3.1.1. Phase transition}\]

It follows easily from (3.1.1) that

\[\text{Theorem 3.1.2. If } \alpha < 1 \text{ then the branching process is subcritical and } \sup_i E|\mathcal{C}_i| < \infty.\]

This conclusion holds for the random graph since the branching process provides an upper bound on cluster growth.

The key to studying the case \( \alpha > 1 \) is a result of about supercritical multitype branching process.

\[\text{Theorem 3.1.3. Kesten and Stigum (1966). If } \alpha > 1 \text{ then there is a one-dimensional random variable } w \text{ so that with probability 1 }\]

\[
Z_n/\alpha^n \to w v
\]
where \( v \) is the right eigenvector. One has either (i) \( E(W|Z_n = e_i) = u_i \) or (ii) \( w = 0 \) with probability one. (i) holds if and only if
\[
E(Z_i^j \log Z_i^j | Z_0 = e_i) < \infty \quad \text{for all } 1 \leq i, j \leq n
\]
This result gives a necessary and sufficient condition for the limit to be nontrivial. In the random graph with finitely many types, \( E(Z_i^j) < \infty \).

**Theorem 3.1.4.** If \( \alpha > 1 \) then the inhomogeneous random graph with finitely many types has a giant component. In the limit the fraction of vertices of type \( k \) in this component is \( \rho_k \), the probability the multitype branching process starting from a single type \( k \) individual does not die out.

**Proof.** This time we cannot reduce the cluster growth process to a random walk, so we will instead run the branching process using a breadth-first search. That is, we first generate the children of the ancestor at time \( t = 1 \), let the number be \( Y_1 \), and write them in an ordered list. We generate the offspring of the first individual in the list, then the second, and so on until we have generated all the children of the first generation. We then order the \( Y_2 \) individuals in the second generation and generate their offspring one at a time. Note that there are two time scales here. \( t \) counts the number of individuals that have given birth, \( m \to Y_m \) gives the size of the \( m \)th generation, which is known at time \( T_m \) on the birth counting time scale.

**Lemma 3.1.5.** Let \( \delta \) be chosen so that \( \bar{\alpha} = \alpha (1 - \delta) > 1 \). At the times \( T_m \) when the growing cluster does not yet contain \( \delta n_k \) vertices in \( V_k \) for some \( k \) the process \( Y_m \) dominates a multitype branching process with mean matrix \( (1 - \delta) \mu_{j,k} \).

Suppose that at the \( t \)th step of the construction we are generating the offspring of a type \( j \) individual. If the growing cluster does not yet contain \( \delta n_k \) vertices in \( V_k \) for any \( k \), we can pick sets \( G_{t,k} \subset V_k \) that have \( |G_{t,k}| = (1 - \delta)n_k \) and \( G_{t,k} \) does not contain any sites of the growing cluster. We allow the vertex \( x \) of type \( j \) to give birth onto each \( y \in G_k \) with probability \( a(j,k)/n \) independent of everything that has happened so far in the construction. This completes the proof of the lemma.

There is no need to do the comparison at time \( t = 1 \) but if we do then \( Y_m \) is exactly multitype branching process with mean matrix \( (1 - \delta)\mu(j,k) \) and in addition the number of individuals of type \( k \) of an individual of type \( j \) is binomial(nk,a(j,k)/n). We say that the process \( Y_m \) survives if the total number of offspring reaches \( n^{2/3} \). Note that if \( n \) is large it follows that the growing cluster does not yet contain \( \delta n_k \) vertices in \( V_k \) for any \( k \).

If the ancestor has type \( k \) then the probability the branching process survives converges to \( \rho_k \) as \( n \to \infty \). Using the limit theorem for supercritical branching processes in Theorem 3.1.3, it is easy to see that if the total progeny reaches \( n^{2/3} \) at time \( M \) then with high probability \( Y_M > n^{0.6} \). Reasoning as in Step 3 in Section 1.3 (and noting that there is an \( \ell \) so that we have \( \mu'(j,k) > 0 \)) it follows that two supercritical branching processes that survive will intersect with high probability.

To complete the proof now we note that there is a \( \beta > 0 \) so that if the multitype branching process has \( \beta \log n \) particle total then the probability that it survives is \( \geq n^{-2} \) independent of the actual collection of types that are present. This follows from part(i) of Thorem 3.1.8. \( \square \)
3.1.2. Computing the survival probability

The basic data for a multitype branching process is

\[ p^j(i_1, \ldots i_m) \]

The probability an individual of type \( j \) gives birth to \( i_1 \) individuals of type 1, \( i_2 \) individuals of type 2, \( \ldots \) \( i_m \) individuals of type \( m \). Here we are following Chapter V of the book of Athreya and Ney (1972) so we will use their notation. Define the multivariate generating function by

\[ f^j(s) = \sum_i p^j(i_1, \ldots i_m)s_1^{i_1} \cdots s_m^{i_m} \quad (3.1.2) \]

where \( s = (s_1, \ldots s_m) \) and combine these into a vector

\[ f(s) = (f^1(s), \ldots f^m(s)) \]

**Example. Random graph.** Since \( N_{j,k} \) are independent binomial\((n_k, a(j, k)/n)\) and the generating function of binomial\((m, p)\) is

\[ \phi(z) = \sum_{\ell=0}^{m} z^\ell \binom{m}{\ell} p^\ell (1-p)^{m-\ell} = (1-p+pz)^m \]

the generating function defined in (3.1.2) is

\[ f^j(s) = \prod_{k=1}^{m} \left( 1 - \frac{a(j, k)}{n} (1-z) \right)^{n_k} \approx \exp \left( -\sum_{k=1}^{m} a(j, k) \pi_k (1-z) \right) \]

By analogy with the one type case we should expect

**Lemma 3.1.6.** Let \( f_k \) be the generating function of \((Z^1_k, \ldots Z^m_k)\). Then \( f_{k+\ell}(s) = f_k(f_\ell(s)) \)

Throughout this section, we will regard facts about multitype branching processes as well-known and not provide proofs. The reader can consult Athreya and Ney for details, although in this case one must go back to Harris’ (1963) book to find the proof.

In the one-type case we have to exclude the situation in which each individual has exactly one offspring. In the multitype case we say that \( Z_n \) is **singular** if this holds.

**Lemma 3.1.7.** If \( Z_n \) is positive regular and nonsingular then

\[ P(Z_n = j \text{ infinitely often}) = 0 \]

for any \( j \geq 0 \) with \( j \neq 0 \).
To prove this it is enough to show that for each \( j \geq 0 \) with \( j \neq 0 \) the probability the process dies out is \( \geq \epsilon_j > 0 \) for then the result follows from Lévy’s 0-1 law, Theorem 4.6.9 in PTE5.

Let \( q^i \) be the probability of extinction for the multitype branching process starting from one individual of type \( i \), and let \( q = (q^1, \ldots, q^m) \).

**Theorem 3.1.8.** Assume \( Z_n \) is positive regular and nonsingular and let \( \alpha \) be the maximum eigenvalue of the mean matrix \( \mu \).

(i) If \( \alpha \leq 1 \) then \( q^i = 1 \) for all \( i \). If \( \alpha > 1 \), \( q^i < 1 \) for all \( i \).

(ii) \( \lim_{n \to \infty} f_n(s) = q \) if \( s \in Q = \{ s : 0 \leq s < 1 \} \).

(iii) \( q \) is the only solution of \( f(s) = s \) in \( Q \)

(iii) follows immediately from (ii). In the one type case we start from \( s = 0 \) and iterate to find \( q \).

### 3.1.3. Subcritical regime

The goal of this section is to show that

**Theorem 3.1.9.** If \( \alpha < 1 \), there are constants \( \theta, \epsilon > 0 \) so that

\[
P(|C_x| \geq k) \leq e^{\theta(1-\epsilon k)}
\]

**Proof.** To prove this we will find the cluster by revealing the neighbors of the vertices one at a time as in Section 1.3. Let \( A_t^i \) be the number of active vertices of type \( i \) at time \( t \). Let \( v_i \) be the right-eigenvector of the mean matrix and let \( Y_t = \sum_i A_t^i v_i \). Let \( r_t \) be the type of the vertex whose neighbors are revealed at time \( t+1 \). Since \( \sum_j \mu(i,j) v_j = \lambda v_i \) if we ignore collisions then \( Y_t \) has negative drift

\[
E(Y_{t+1} - Y_t | r_t = i) = (\lambda - 1)v_i
\]

The assumption of no collisions is valid since with high probability the number of sites involved will be \( O(\log n) \). \( Y_t \) is not a random walk but it only has finitely many different types of jumps, so we can find an exponential martingale. Let \( F_t \) be the information known at time \( t \) and let \( \epsilon > 0 \). On \( \{r_t = i\} \)

\[
E(e^{\theta[Y_t+1]} | F_t) = e^{\theta[Y_t]} E \left( \exp \left( \epsilon - \theta v_i + \theta \sum_j N_{i,j} v_j \right) \mid F_t \right)
\]  

(3.1.3) **inhsub1**

Ignoring the \( e^{\epsilon - \theta v_i} \), the conditional expectation on the right-hand side is

\[
\prod_j \left( 1 - \frac{a(i,j)}{n} + \frac{a(i,j)}{n} e^{\theta v_j} \right)^{n_j} \approx \exp \left( \mu(i,j) [e^{\theta v_j} - 1] \right)
\]  

(3.1.4) **inhsub2**
since $\mu(i, j) = a(i, j)\pi_j$.

The inequality $e^{\theta v_i} - 1 \geq \theta v_i$ goes the wrong way for finding a supermartingale so we need a complimentary upper bound. The next result is routine and the proof is easy but we include the details for completeness.

**Lemma 3.1.10.** Let $\eta > 0$. If $\delta$ is small enough then for all $0 \leq x \leq \delta$, $e^x - 1 \leq (1 + \eta)x$.

**Proof.** Using the power series expansion for $e^x$ we see that if $0 \leq x \leq \delta$

$$e^x - 1 = \sum_{m=1}^{\infty} \frac{x^m}{m!} \leq x + x \left( \frac{\delta}{2!} + \frac{\delta^2}{3!} + \cdots \right) \leq x \left( 1 + \frac{\delta}{2(1 - \delta)} \right)$$

since $k! \geq 2$ for $k \geq 2$. Choosing $\delta$ small proves the result. \qed

Using (3.1.3) and (3.1.4) with the Lemma, we see that if $\theta$ is chosen small enough so that $0 \leq \theta v_i \leq \delta$ for all $i$ then on $\{r_t = i\}$

$$E(e^{\theta[Y_{t+1}+\epsilon(t+1)]}\mid \mathcal{F}_t) \leq e^{\theta[Y_{t+\epsilon}] \exp \left( \epsilon - \theta v_i + \theta(1 + \eta) \sum_j \mu(i, j)v_j \right)}$$

The last quantity on the right

$$\epsilon - \theta v_i + \theta(1 + \eta) \sum_j \mu(i, j)v_j = \epsilon - \theta v_i(1 - \lambda(1 + \eta)).$$

If we pick $\eta$ small enough so that $\lambda(1+\epsilon) < 1$ and then $\epsilon$ small $\exp(\theta[Y_{t+\epsilon}]$) is a supermartingale. Let $\tau$ be time that the exploration process terminates and recall that if we start with one active site $x$ then $\tau = |C_x|$. Using stopping theorem for nonnegative supermartingales, Theorem 4.8.4 in PTE5, we have

$$e^\theta \geq E e^{\theta \tau} \geq e^{\theta \epsilon k} P(|C_x| \geq k)$$

which proves the desired result. \qed
3.2 Motivating Examples

The concept of an inhomogeneous graphs, as we define it here, was introduced in a paper by Bollobás, Janson and Riordan (2007) that fills an entire issue of Random Structures and Algorithms. To motivate the machinery that needs to be developed, we will first show that it covers a number of interesting examples.

3.2.1. Uniformly grown random graphs (Dubins’ model)

Kalikow and Weiss (1988) considered a random graph on the positive integers \( \mathbb{N} \) in which there is an edge between \( i \) and \( j \) with probability

\[
p_{ij} = \left( \frac{\lambda}{\max\{i, j\}} \right) \wedge 1
\]

where \( a \wedge b = \min\{a, b\} \). The existence of an infinite component is trivial: for any \( \lambda \in (0, 1] \), \( p_{1,j} = \lambda/j \), so the component containing 1, \( C_1 \) is always infinite. Because of this, they considered the question: Is \( \mathbb{N} \) connected? They showed that the answer was yes if \( \lambda > 1 \) and no if \( \lambda < 1/4 \). Dubins had earlier conjectured that the critical value \( \lambda_c = 1 \). Shepp (1989) later proved that the answer was yes if \( \lambda > 1/4 \) and no if \( \lambda \leq 1/4 \), so \( \lambda_c = 1/4 \).

Callaway, Hopcroft, Kleinberg, Newman, and Strogatz (2001) introduced the following simple version of a randomly grown graph. Start with \( G_1 = \{1\} \) with no edges. At each time \( n \geq 2 \), we add one vertex labelled \( n \) and with probability \( \delta \) add one edge between two randomly chosen vertices. Note that the newly added vertex is not necessarily an endpoint of the added edge and when \( n \) is large, it is likely not to be.

CHKNS began their analysis by examining \( N_k(t) = \) the expected number of components of size \( k \) at time \( t \). Ignoring terms of \( O(1/t^2) \), which come from picking the same cluster twice:

\[
N_1(t + 1) = N_1(t) + 1 - 2\delta \frac{N_1(t)}{t}
\]

\[
N_k(t + 1) = N_k(t) - 2\delta \frac{kN_k(t)}{t} + \delta \sum_{j=1}^{k-1} jN_j(t) \cdot \frac{(k-j)N_{k-j}(t)}{t}
\]

(3.2.1)  

(3.2.2)  

To explain the first equation, note that at each discrete time \( t \) one new vertex is added, and a given isolated vertex becomes the endpoint of an added edge with probability \( \approx 2\delta/t \). For the second equation, note that the probability an edge connects to a given cluster of size is \( \approx 2\delta k/t \), while the second term corresponds to mergers of clusters of size \( j \) and \( k-j \). There is no factor of 2 in the last term since we sum from 1 to \( k-1 \).

**Theorem 3.2.1.** As \( t \to \infty \), \( N_k(t)/t \to a_k \) where \( a_1 = 1/(1 + 2\delta) \) and for \( k \geq 2 \)

\[
a_k = \frac{\delta}{1 + 2\delta k} \sum_{j=1}^{k-1} ja_j \cdot (k-j)a_{k-j}
\]
To solve for the $a_k$, which gives the limiting number of clusters of size $k$ per site, CHKNS used generating functions. Let $h(x) = \sum_{k=1}^{\infty} x^k a_k$ and $g(x) = \sum_{k=1}^{\infty} x^k k a_k$. Multiplying the formulas for $a_k$ in Theorem 3.2.1 by $(1+2\delta k)x^k$, recalling the formula for $a_1$ is different from the others, and summing gives

$$h(x) + 2\delta g(x) = x + \delta g^2(x) \quad (3.2.3)$$

Since $h'(x) = g(x)/x$ differentiating (3.2.3) gives $g(x)/x + 2\delta g'(x) = 1 + 2\delta g(x)g'(x)$. Rearranging we have $2\delta g'(x)(1 - g(x)) = 1 - g(x)/x$

$$g'(x) = \frac{1}{2\delta x} \cdot \frac{x - g(x)}{1 - g(x)} \quad (3.2.4)$$

Let $b_k = k a_k$ be the number of vertices that belong to clusters of size $k$. $g(1) = \sum_{k=1}^{\infty} b_k$ gives the fraction of vertices that belong to finite components. $1 - g(1)$ gives the fraction of sites that belong to clusters whose size grows in time. Even though it is not known that the missing mass in the limit belongs to a single cluster, it is common to call $1 - g(1)$ the fraction of sites that belong to the giant component. The next result gives the mean size of finite components. Here and throughout the motivating examples we state results without proving them.

![Figure 3.1: Graph of the result in Theorem 3.2.3.](image-url)
3.2. MOTIVATING EXAMPLES

Using the differential equation (3.2.4) one can show

**Lemma 3.2.2.** (i) If \( g(1) < 1 \) then \( \sum_{k=1}^{\infty} kb_k = g'(1) = 1/2\delta \).

(ii) If \( g(1) = 1 \) then \( g'(1) = (1 - \sqrt{1 - 8\delta})/4\delta \).

**Theorem 3.2.3.** The critical value \( \delta_c = \sup\{\delta : g(1) = 1\} = 1/8 \) and hence

\[
\sum_k kb_k = \begin{cases} 
    (1 - \sqrt{1 - 8\delta})/4\delta & \delta \leq 1/8 \\
    1/2\delta & \delta > 1/8
\end{cases}
\]

Note that this implies that the mean cluster size \( g'(1) \) is always finite, but is discontinuous at \( \delta = 1/8 \), since the value there is 2 but the limit for \( \delta \downarrow 1/8 \) is 4.

**Durrett (2003)**

modified the model so that a Poisson mean \( \delta \) number of vertices are added at each step. This version is nicer since in the Poisson case if we let \( A_{i,j,k} \) be the event no \((i,j)\) edge is added at time \( k \) then \( P(A_{i,j,k}) = \exp\left(-\delta/(\frac{k}{2})\right) \) for \( i < j \leq k \) and these events are independent.

\[
P(\cap_{k=j}^{n} A_{i,j,k}) = \prod_{k=j}^{n} \exp\left(-\frac{2\delta}{k(k-1)}\right) = \exp\left(-2\delta \left(\frac{1}{j-1} - \frac{1}{n}\right)\right) \geq 1 - 2\delta \left(\frac{1}{j-1} - \frac{1}{n}\right) \quad \text{#1}
\]

The last formula is not simple, so he also consider two approximations

\[
a \approx 1 - 2\delta \left(\frac{1}{j-1} - \frac{1}{n}\right) \quad \text{#2}
\]

\[
a \approx 1 - \frac{2\delta}{j} \quad \text{#3}
\]

Referring to these three models by their numbers, he showed that

**Theorem 3.2.4.** In models #1, #2, or #3, the critical value for the existence of a component of order \( n \) is \( \delta_c = 1/8 \).

In addition Durrett (2003) obtained some results about the decay of connection probabilities at the critical value and in the subcritical regime. See Section 7.3 in RGD.

**Bollobás, Janson, and Riordan (2005)**

were not aware of Durrett’s work. They did an independent analysis of the CHKNS model, proving a remarkable result about the supercritical phase: the size of the giant component is infinitely differentiable at the critical value. To put their work into context we need to describe previous work on the size of the giant component.
To investigate the size of the giant component, CHKNS integrated the differential equation (3.2.4) near $\delta = 1/8$. Letting $S(\delta) = 1 - g(1)$ the fraction of vertices in the infinite component they plotted $\log(-\log S)$ vs $\log(\delta - 1/8)$ and concluded that

$$S(\delta) \sim \exp(-\alpha(\delta - 1/8)^{-\beta})$$

where $\alpha = 1.132 \pm 0.008$ and $\beta = 0.499 \pm 0.001$. Based on this they conjectured that $\beta = 1/2$. Inspired by their conjecture Dorogovstev, Mendes, and Samukhin (2001) calculated (using results that were not completely rigorous but turned out to be very accurate) that as $\delta \downarrow 1/8$,

$$S \equiv 1 - g(1) \approx c \exp(-\pi/\sqrt{8\delta} - 1) \quad \text{(3.2.5)}$$

This result shows $\beta = 1/2$ and $\alpha = \pi/\sqrt{8} = 1.1107$. For a look at the awesome power of physics to study problems of this type see the survey article by Dorogovstev and Mendes (2002)

BJR(2005) used $c = 2\delta$ as their parameter. They showed that if $\eta > 0$

$$\exp\left(-\frac{\pi + \eta}{2\sqrt{\epsilon}}\right) \leq C_1(G_n(1/4 + \epsilon)) \leq \exp\left(-\frac{1 - \eta}{2\sqrt{\epsilon}}\right)$$

Riordan (2005)

improved the result to give the sharp constant.

$$C_1(G_n(1/4 + \epsilon)) \approx \exp\left(-\frac{\pi}{2\sqrt{\epsilon}} + O(\log(1/\epsilon))\right)$$

which agrees with (3.2.5)

### 3.2.2. Preferential attachment

Barabási and Albert (1999)

introduced this model. Intuitively, at every time step, we add a new vertex with $m$ edges that link the new vertex to $m$ different vertices already present in the system. To incorporate preferential attachment, we assume that the probability $\Pi$ that a new vertex will be connected to a vertex $i$ depends on the connectivity of that vertex.

To have a precise definition, we will suppose that the process starts at time 1 with two vertices linked by $m$ parallel edges, so that the total degree at any time $t$ is $2mt$. When we add a new vertex we will add edges one a time, with the second and subsequent edges doing preferential attachment using the updated degrees. This scheme has the desirable property that a graph of size $n$ for a general $m$ can be obtained by running the $m = 1$ model for $nm$ steps and then collapsing vertices $km, km - 1, \ldots, (k - 1)m + 1$ to make vertex $k$. 
They did simulations that suggested a power law distribution with power $\gamma = 2.9 \pm 0.1$ and gave the following argument for $\gamma = 3$. If we consider the degree of $i$, $k_i$, to be a continuous variable then

$$\frac{\partial k_i}{\partial t} = m \cdot \frac{k_i}{\sum_j k_j} = \frac{k_i}{2t}$$

since $\sum_j k_j = 2mt$. The solution to this differential equation is

$$k_i(t) = m(t/t_i)^{1/2}$$

where $t_i$ is the time the vertex was introduced. From this we see that

$$P(k_i(t) > k) = P(t_i < tm^2/k^2) = m^2/k^2$$

since vertices are added uniformly on $[0,t]$. Differentiating

$$P(k_i(t) = k) = 2m^2/k^3$$

Dorogovstev, Mendes, and Samukhin (2000) took a different approach, using what they call the “master equation,” which can be used to obtain rigorous asymptotics for the mean number of vertices of degree $k$, $N(k,t)$. By considering what happens on one step, in which vertices of degree $k$ can be created from those of degree $k - 1$ or lost by becoming degree $k + 1$, we see that

$$N(k,t + 1) - N(k,t) = \frac{m(k - 1)}{2mt} N(k - 1,t) - \frac{mk}{2mt} N(k,t) + \delta_{k,m} \quad (3.2.6)$$

Here we ignore the possibility of more than one edge being attached to one vertex, and the updating of edges that occurs as the $m$ edges are added. The last term takes care of the fact that when $k = m$ we add one vertex of degree $m$ at each time. To make the equation correct for $k = m$ we suppose $N(j,t) = 0$ for $j < m$.

Letting $p_k = \lim_{t \to \infty} N(k,t)/t$, and using (3.2.6) we conclude that

$$p_m = 2/(m + 2)$$

$$p_k = \frac{(k - 1)p_{k-1}}{2} \cdot \frac{2}{k + 2} = p_{k-1} \frac{k - 1}{k + 2}$$

The solution to this recursion is

$$p_k = \frac{2m(m + 1)}{k(k + 1)(k + 2)} \quad (3.2.7)$$

Bollobás and Riordan (2004) have performed a rigorous analysis of the diameter of the graphs produced by the Barabási-Albert model. First, consider the case $m = 1$. They inductively define a sequence of directed
random graphs $G_t^i$ on $\{i : 1 \leq i \leq t\}$. Start with $G_1^1$ the graph with one vertex and one loop. Given $G_{t-1}^i$ form $G_t^i$ by adding the vertex $t$ together with a directed edge from $t$ to $I$ where $I$ is chosen randomly with

$$P(I = i) = \begin{cases} 
\frac{d_i^{t-1}}{(2t-1)} & 1 \leq i \leq t - 1 \\
\frac{1}{(2t-1)} & i = t
\end{cases}$$

(3.2.8) \text{BRchoice}

where $d_i^{t-1}$ is the degree of $i$ in $G_{t-1}^i$. In words, we consider the outgoing edge from $t$ when we consider where to attach the other end of the edge. Note that each vertex will have out degree 1. To extend the definition to $m > 1$, use the definition above to define random graphs $G_{mt}^1$ on $\{v_i : 1 \leq i \leq mt\}$ then combine the vertices $v_{(k-1)m+1}, \ldots v_{km}$ to make vertex $k$.

Their main result is:

**Theorem 3.2.5.** Let $m \geq 2$ and $\epsilon > 0$. Then with probability tending to 1, $G_m^m$ is connected and

$$(1 - \epsilon) \log n/(\log \log n)) \leq \text{diameter}(G_m^m) \leq (1 + \epsilon) \log n/(\log \log n))$$

The case $m = 1$, in which the graph is a tree, is excluded because the upper bound is false in this case.

In order to prove the lower bound they consider $G_1^N$ with $N = nm$ and compare $G_1^N$ with a random graph in which an edge from $i$ to $j$ is present with probability $c/\sqrt{ij}$. Let $g_j$ be the vertex to which $j$ sends an edge when it is added to the graph.

**Lemma 3.2.6.** (a) If $1 \leq i < j$ then $P(g_j = i) \leq C_1(ij)^{-1/2}$.
(b) If $1 \leq i < j < k$ then $P(g_j = i, g_k = i) \leq C_2(i^{-1}(jk))^{-1/2}$.

This proof motivates the consideration of the inhomogeneous graph with

$$p(i, j) = c/\sqrt{ij}$$

(3.2.9) \text{sqrtmodel}

which is often called the mean-field version of preferential attachment because edges are independent. There is a giant component in this graph for all $c > 0$.

**Riordan (2005)**

has shown for the square root model defined in (3.2.9) that there is a function $f(c)$ so that with high probability the largest component has $(f(c) + o(1))n$ vertices as $n \to \infty$ where

$$f(c) \sim 2e^{1-\gamma} \exp(-1/(2c))$$

(3.2.10) \text{sqrtgiant}
3.2.3. Chung-Lu model

In this model weights $w_i$ are assigned to vertices and we define the probability of an edge between $i$ and $j$ to be

$$p(i, j) = \frac{w_i w_j}{\sum_k w_k} \wedge 1$$

(3.2.11) \textbf{CLdef}

The weight $w_i$ is the expected degree of $i$ since (ignoring the $\wedge 1$ and including the $j = i$ (self-loop) term in the sum.)

$$\sum_j w_i w_j = w_i$$

If we let

$$w_i = (i/Cn)^{-1/(\gamma - 1)}$$

(3.2.12) \textbf{powerCL}

then $w_i > w$ if and only if $i/n < Cw^{-(\gamma - 1)}$ so the graph has a power law degree distribution. $P(w_i > w) \sim Cw^{-(\gamma - 1)}$. If $\gamma > 2$ then $Ew_i < \infty$ so $\sum_k w_k \sim \mu n$.

To get the Chung-Lu model we take $a = 1/(\gamma - 1)$, $\psi(x) = x^{-a}$ and $\kappa(x, y) = c\psi(x)\psi(y)$

In this case

$$p(i, j) = \kappa(i/n, j/n)/n = \frac{c(i/n)^{-a}(j/n)^{-a}}{n}$$

Chung and Lu proved a number of results about their model including the following.

\textbf{Theorem 3.2.7.} Consider Chung and Lu’s power law graphs with $2 < \gamma < 3$. Then the distance between two randomly chosen vertices in the giant component, $H_n$ is asymptotically at most

$$(2 + o(1)) \log \log n/(−\log(\gamma − 2)).$$

The proof uses the following interesting result, which shows an advantage of the inhomogeneous random graph formulation.

\textbf{Lemma 3.2.8.} If $S \cap T = \emptyset$ and $\text{Vol}(S)\text{Vol}(T) \geq c\text{Vol}(G)$ then the distance from $S$ to $T$ satisfies $P(d(S,T) > 1) \leq e^{-c}$.

\textbf{Proof.} Since $1 - x \leq e^{-x}$

$$P(d(S,T) > 1) = \prod_{i \in S, j \in T} 1 - \frac{w_i w_j}{\text{Vol}(G)} \leq \exp \left( - \sum_{i \in S, j \in T} \frac{w_i w_j}{\text{Vol}(G)} \right) \leq e^{-c}$$

$\square$
3.3 Welcome to the machine

The description of the framework used in BJR(2007) will require a number of definitions. We will restrict ourselves to a special case that is enough to cover the examples we are interested in.

A **ground space** is a pair \((S, \mu)\) where \(S\) is a separable metric space and \(\mu\) is a Borel probability measure. For us this will usually be \(S = (0, 1]\) and \(\mu = \text{Lebesgue measure, i.e., } d\mu(x) = dx\).

A **vertex space** \(V\) is a triple \((S, \mu, \{x^n\}_{n \geq 1})\) where \((S, \mu)\) is a ground space and \(x^n = (x^n_1, \ldots, x^n_n)\) is a random sequence of points in \(S\) so that

\[
\nu_n(A) = \frac{\# \{ i \leq n : x^n_i \in A \}}{n} \to \mu(A) \quad (3.3.1)
\]

for every \(\mu\)-continuity set, i.e., \(\mu(\partial A) = 0\) where \(\partial A\) is the boundary of \(A\). In most of our examples \(x^n_i = i/n\). We will not consider **generalized vertex spaces** where the number of vertices is random.

A **kernel** \(\kappa\) on a ground space \((S, \mu)\) is a symmetric non-negative Borel measurable function on \(S \times S\).

Given a vertex space and a kernel, we can define a random graph \(G^V(n, \kappa)\) in which an edge from \(i\) to \(j\) is present with probability

\[
p^n_{ij} = \frac{\kappa(x^n_i, x^n_j)}{n} \land 1
\]

To avoid taking the minimum with 1 one can define

\[
p^n_{ij} = 1 - \exp(-\kappa(x^n_i, x^n_j)/n)
\]

A kernel \(\kappa\) is **graphical** if (i) \(\kappa\) is continuous almost everywhere on \(S \times S\), (ii) \(\kappa \in L^1(S \times S, \mu \times \mu)\) and (iii) if \(e(G)\) is the number of edges in \(G\)

\[
(iii) \quad \frac{1}{n} Ee(G^V(n, \kappa)) \to \frac{1}{2} \iint \kappa(x, y) \, d\mu(x) \, d\mu(y)
\]

(ii) and (iii) imply that the expected number of edges is \(O(n)\).

If \(\kappa \in L^1(S \times S, \mu \times \mu)\) then \(\int_S \kappa(x, y) \, d\mu(y) < \infty\) for \(\mu\)-a.e. \(x\). It is convenient to assume that

\[
\int_S \kappa(x, y) \, d\mu(y) < \infty \quad \text{for all } x \in S \quad (3.3.2)
\]

It seems to me that the main reason for allowing \(\kappa\) to be discontinuous is to allow for finitely many types example. Since we will consider only examples on \((0, 1]\) we will restrict our attention to kernels that are **piecewise continuous**. That is there are \(0 = a_0 < a_1 < a_2 < \)
... $a_k = 1$ so that $\kappa$ is continuous on each $(a_{i-1}, a_i)$. In addition we require that on each interval $(a_{i-1}, a_i)$ $\kappa$ converges to a finite limit as either endpoint is approached, unless $a_{i-1} = 0$ in which case we can have $\kappa \to \infty$.

It is useful to have some concrete examples to consider. In all these cases $\mathcal{S} = (0, 1]$, $\mu$ is Lebesgue measure and $x_i^n = i/n$.

Example 1. Finitely many types. For $1 \leq k \leq K$ we have intervals $I_k = (a_{k-1}, a_k]$ with $0 = a_0 < a_1 < \ldots < a_K = 1$. For $x \in I_k$, $y \in I_\ell$, $\kappa(x, y) = a(k, \ell)$.

Example 2. Square root model. $\kappa(x, y) = c/\sqrt{xy}$. In this case $\kappa \not\in L^1$. As we will see this implies that the critical value $c_c = 0$.

Example 3. Dubins model. $\kappa(x, y) = c/\max\{x, y\}$

Example 4. Durrett and Kesten (1990) considered a model with $\kappa(x, y) = c/(x^p + y^p)^{1/p}$. They suppose $p > 0$. Here we will suppose $p \geq 1$ so that condition (ii) holds. Letting $p \to \infty$ gives the Dubins model.

Example 5. Chung and Lu (2002). $\kappa(x, y) = c\psi(x)\psi(y)$ with $\psi(x) = x^{-1/(\gamma-1)}$.

3.3.1. Phase transition

To begin we define an integral operator

$$(T_\kappa f)(x) = \int_{\mathcal{S}} \kappa(x, y)f(y) \, d\mu(y)$$

The norm $\|T_\kappa\| = \sup\{\|T_\kappa f\|_2 : f \geq 0, \|f\|_2 \leq 1\}$. In the finite type case this operator is essentially the mean matrix, and the norm is the same as the maximum eigenvalue. To be precise if we let $g(x) = w_k$ for $x \in I_k$ (defined in Example 1 above) then

$$(T_\kappa g)(x) = \sum_k \mu(j, k)w_k \quad \text{for } x \in I_j$$

In addition we define

$$\Phi_\kappa f = 1 - \exp(-T_\kappa f)$$

The survival probability $\rho_\kappa(x)$ is the maximum fixed point of $\Phi$. Recall that in the Erdős-Rényi and configuration models the extinction probability is defined in terms of the smallest fixed point of generating functions. In the Erdős-Rényi case when the mean degree is $\lambda > 1$ the extinction probability satisfies

$$z = \exp(-\lambda(1 - z))$$

so $\rho = 1 - z$ satisfies $\rho = 1 - \exp(-\lambda \rho)$.

To state the next result we need a definition. $\kappa$ is reducible if there is an $A$ with $0 < \mu(A) \leq \mu(\mathcal{S})$ so that $\kappa(x, y) = 0$ on $A \times (\mathcal{S} - A)$. Otherwise it is irreducible.
\section*{Chapter 3. Inhomogeneous Random Graphs}

**Theorem 3.3.1.** (i) If $\|T_\kappa\| < 1$ then the largest component $C_1$ is $o(n)$ while if $\|T_\kappa\| > 1$ it is $\Theta(n)$. (ii) If $\kappa$ is irreducible then

$$(1/n)|C_1| \to \rho = \int_0^1 \rho_\kappa(x) d\mu(x).$$

The reasoning is similar to that for Erdős-Rényi. $x$ is part of the giant component if the branching process started at $x$ does not die out.

Turning to our examples, the **rank 1 case** $\kappa(x, y) = c\psi(x)\psi(y)$ is the easiest. In this case the operator $T$ has

$$T_\kappa f = e \left( \int f \psi d\mu \right) \cdot \psi$$

Since $T$ maps all functions to multiples of $\psi$, $\psi$ is the only possible eigenvector. It has eigenvalue $c \int \psi(x)^2 dx$ and critical value

$$c_c = 1/ \int \psi(x)^2 dx.$$ 

In the square root model $\psi(x) = cx^{-1/2}$ which has $\int \psi^2 dx = \infty$, so $c_c = 0$. 

In the **Chung-Lu model**, $\psi(x) = x^{-1/(\gamma-1)}$ so $c_c = 0$ if $\gamma \leq 3$ while it is $>0$ if $\gamma > 3$.

**Dubins model.** BJR observe in Section 16.1 that

$$U : f \to e^{-x/2} f(e^{-x})$$

is an isometry from $L^2(0, 1)$ onto $L^2(0, \infty)$. The operator $\tilde{T}_\kappa = U T_\kappa U^{-1}$ has kernel

$$\tilde{\kappa}(x, y) = e^{-x/2} \kappa(e^{-x}, e^{-y}) e^{-y/2} = e^{-x/2 - y/2 + x \wedge y} = e^{-|x-y|/2}$$

Thus $\tilde{T}_\kappa$ is the restriction to $(0, \infty)$ of the convolution with $h(x) = ce^{-|x|/2}$. “Taking Fourier transforms we find $\|T_\kappa\| = 4c$.” Since it is not hard to find the critical value of the model directly, will not go into the details.

**Durrett and Kesten (1990).** $\kappa(x, y) = c/(x^p + y^p)^{1/p}$ is homogeneous of degree $-1$, i.e., $\kappa(ax, ay) = a^{-1}$ so

$$p(i, j) = \kappa(i/n, j/n)/n = c/(i^p + j^p)^{1/p}$$

independent of $n$. In Dubins’ model $\kappa(x, y)$ is homogeneous of degree $-1$.

The main result in DK(1990) is

**Theorem.** Suppose $p(i, j) = \min\{\lambda h(i, j), 1\}$ where the function $h : [0, \infty)^2 - \{0\} \to (0, \infty)$ is symmetric $h(x, y) = h(y, x)$, homogeneous of degree $-1$, and continuous and strictly positive on $\{(x, y) : x + y = 1\}$. Then the critical value for $\lambda$ is

$$\lambda_c = \left( \int_0^1 \frac{h(u, 1-u)}{\sqrt{u(1-u)}} du \right)^{-1} = \left( \int_0^1 \frac{h(1, y)}{\sqrt{y}} dy \right)^{-1}$$
As in the work of Kalikow and Weiss (1988) mentioned in the discussion of Dubins model, DK(1990) was concern with the threshold for the positive integers $N$ to be connected with positive probability. However if one assumes that $h$ is decreasing in each of its arguments which holds for $h(x, y) = 1/(x^p + y^p)^{1/p}$ and writes $H_{K,M}$ for the submatrix of $h(i, j)$ with $K \leq i, j, \leq M$ then
\[
\lim_{M \to \infty} \rho(H_{1,M}) = \int_0^1 \frac{h(u, 1-u)}{\sqrt{u(1-u)}} \, du
\]
so $\lambda_c$ is also the critical value for a giant component in $\{1, 2, \ldots n\}$. To compute the critical value DK(1990) show that as Shepp(1989) suggested $w_k = k^{-1/2}$ is a good approximation of the eigenvector associated with the largest eigenvalue $\rho$ so
\[
\lambda_c(p) = p\Gamma(1/p) [\Gamma(1/2p)]^{-2}
\]
where $\Gamma$ is the usual Gamma function.

### 3.3.2. Degree distribution

In the special case $S = (0, 1]$ with $\mu$ Lebesgue measure, we define
\[
\lambda(x) = \int_0^1 h(x, y) \, dy
\]
to be the **expected degree** of $x$.

**Theorem 3.13 in BJR.** Writing $Z_k$ for the number of vertices of degree $k$,
\[
Z_k/n \to P(D = k) = \int_0^1 e^{-\lambda(x)} \frac{\lambda(x)^k}{k!} \, dx
\]
In words, the degree $D$ of a randomly chosen vertex has the mixed Poisson distribution
\[
\int_0^1 \text{Poisson}(\lambda(x)) \, dx.
\]
Let $\Lambda$ be the expected degree of a randomly chosen vertex. That is, $\Lambda = \lambda(U)$ where $U$ is uniform on $(0, 1]$.

**Corollary 3.13 in BJR.** Suppose that
\[
P(\Lambda > t) = \mu \{ x : \lambda(x) > t \} \sim at^{-\alpha}
\]
for some $a > 0$ and $\alpha > 1$. Let $Z_{\geq k}$ be the number of vertices with degree $\geq k$ then
\[
Z_{\geq k}/n \to P(D \geq k) \sim ak^{-\alpha}
\]
Proof. By Theorem 3.13 it suffices to show that \( P(D \geq k) \sim ak^{-\alpha} \). To do this we note that if \( Z = \text{Poisson}(\lambda) \) then \((Z - \lambda)/\lambda^{1/2}\) converges to a Normal distribution as \( \lambda \to \infty \) so if \( 1/2 < b < 1 \)

\[
P(\lambda - \lambda^b < \text{Poisson}(\lambda) < \lambda + \lambda^b) \to 1
\]

Using this we conclude that if \( k \) is large and \( \epsilon > 0 \)

\[
P(\text{Poisson}(\lambda(x)) \geq k) \approx \begin{cases} 
1 & \text{if } \lambda(x) \geq (1 + \epsilon)K \\
0 & \text{if } \lambda(x) \leq (1 - \epsilon)K 
\end{cases}
\]

If we suppose for simplicity that \( \lambda(x) \) is decreasing we have

\[
P(D \geq k) \approx \mu\{x : \lambda(x) \geq k\}
\]

which gives the desired result. 

In a rank-one model

\[
\lambda(x) = c\psi(x) \int_0^1 \psi(y) \, dy
\]

If \( \psi(x) = x^{-a} \) and \( a < 1 \) then

\[
\lambda(x) = cx^{-a} \int_0^1 y^{-a} \, dy = cx^{-a}/(1 - a)
\]

Letting \( B = c/(1 - a) \) we have \( \lambda(x) > t \) if \( Bx^a > t \) or \( x < (t/B)^{-1/a} \). If \( t > B \) then

\[
P(\Lambda > t) = (t/B)^{-1/a}
\]

In the Chung-Lu model \( a = 1/(\gamma - 1) \), so \( P(\Lambda > t) \sim Ct^{-\gamma+1} \) so the degree distribution has \( p_k \sim C^\gamma k^{-\gamma} \).
3.4 Results for the survival probability

This is done by comparing with a multitype branching process. Here we follow Section 5 of Riordan (2005). To define the branching process we declare that the children of $y$ are a Poisson process with mean measure $\kappa(y, z) \, dz$ on $(0, 1]$. That is, if we let $N(A)$ be the number of points in $A$ then

(i) If $A_1, \ldots, A_n$ are disjoint $N(A_1), \ldots, N(A_k)$ are independent

(ii) $N(A)$ has Poisson $\left( \int_A \kappa(y, z) \, dz \right)$ distribution.

The next result will make this definition more explicit.

**Lemma 3.4.1.** If a particle in $X_k$ has type $a$ and $m$ children they will have joint distribution

$$f_m(a, b_1, \ldots b_m) = \exp \left( - \int_0^1 \kappa(a, b) \, db \right) \frac{1}{m!} \prod_{i=1}^m \kappa(a, b_i)$$

*Proof.* To see this suppose we have points $1/n, 2/n, \ldots 1$ and a random graph with

$$p(i, j) = \kappa(i/n, j/n) / n$$

The probability that $j$ connects to $k_1, \ldots k_m$ and to no other vertices is

$$= \prod_{k \neq k_i} (1 - \kappa(j/n, k/n) / n) \prod_{i=1}^m \kappa(j/n, k_i/n) \cdot \frac{1}{n}$$

Letting $a = j/n$ nd $b_i = k_i/n$, the first term is

$$\approx \exp \left( - \int_0^1 g(\alpha, \beta) \, d\beta \right)$$

If we think of $b_i$ as an interval of length $1/n$ and $g(a, b)$ as a density function the second term is

$$\approx \prod_{i=1}^m g(a, b_i)$$

If we get rid of the ordering $\beta_1 < \cdots \beta_m$ we need to divide by $m!$.

Our next task is to define the success probability. In Riordan (2005)'s approach we let

$$F(g)(a) = \sum_{m \geq 0} \int_{b_1 \ldots b_m \in S} f_m(a, b_1, \ldots b_m) \left( 1 - \prod_{i=1}^m (1 - g(b_i)) \right) \, db_m \cdots db_1$$

If we assign to each child $b$ an event called success that has probability $g(b)$ then $F(g)(a)$ is probability of success of at least one child of $a$. If we let $S_\infty(b)$ be the probability that the branching process starting with one individual of type $b$ does not die out then $S_\infty = F(S_\infty)$.

To develop the properties of $S_\infty$ we need we will follows Section 5 in BJR who write things using a different notation.
**Lemma 3.4.2.** Consider the random offspring of a single particle of type $x$ and let $N$ be the total number of children. If $0 \leq g \leq 1$ is measurable

$$(\Phi_\kappa g)(x) \equiv 1 - E \prod_{i=1}^{N} (1 - g(\xi_i)) = 1 - e^{-(T_\kappa g)(x)}$$ \hspace{1cm} (3.4.1)\hspace{1cm}$$

where $\equiv$ indicates that the first equality is a definition

**Proof.** Let $d\nu(y) = \kappa(x,y)\,dy$ and $\nu' = \nu/\nu(S)$ Given $N = n$, if we list the locations in random order $\xi_1, \xi_2, \ldots \xi_n$ the locations are independent with distribution $\nu'$ so

$$\prod_{i=1}^{n} E(1 - g(\xi_i)) = (E[1 - g(\xi_1)])^n = \left(1 - \int_0^{1} g(y) \, d\nu'(y)\right)^n = \left(1 - \frac{(T_\kappa g)(x)}{\nu(S)}\right)^n$$

Using $P(N = n) = e^{-\nu(S)\nu(S)/n!}$ taking expectation, and writing $h(x) = (T_\kappa g)(x)$ to simplify the formulas gives

$$E \prod_{i=1}^{N} (1 - g(\xi_i)) = \sum_{n=0}^{\infty} e^{-\nu(S)\nu(S)/n!} \cdot \left(1 - \frac{h(x)}{\nu(S)}\right)^n$$

$$= e^{-h(x)} \sum_{n=0}^{\infty} e^{-\nu(S)+h(x)} \frac{[\nu(S) - h(x)]^n}{n!} = e^{-h(x)}$$

which proves the second equality. \hfill \Box

**Lemma 3.4.3.** (i) For $m \geq 0$ the probability that a particle of type $x$ has descendants in at least $m$ further generations is $(\Phi_m^1)(x)$

(ii) As $m \to \infty$, $(\Phi_m^1)(x)$ decreases to a limit we call $\rho_\kappa(x)$

(iii) $\rho_\kappa$ satisfies $\Phi_\kappa \rho_\kappa = \rho_\kappa$

(iv) If $f(x)$ is any other solution $\rho_\kappa(x) \geq f(x)$.

**Proof.** Let $g_m(x)$ be the probability that a particle of type $x$ has descendants in at least $m$ further generations. If we let $g = g_m$ in (3.4.1) then $E \prod_{i=1}^{N} (1 - g_m(\xi_i))$ is the probability that none of the children of $x$ has descendants in the $m+1$ generation. This is $1 - \Phi_\kappa g_m$ and the result follows by induction.

(ii) follows immediately from (i). To prove (iii) we note that

$$\Phi_m^{m+1} 1 = \Phi_\kappa (\Phi_m^1)$$

then let $m \to \infty$ and use the dominated convergence theorem. To prove (iv) we note that if $1 \geq f(x)$ and $f$ is a solution of $\Phi_\kappa f = f$ then it follows by induction that $\Phi_m^1 \geq f$. Letting $m \to \infty$ gives the desired result. \hfill \Box
Lemma 3.4.4. Suppose that $f \geq 0$ with $f = \Phi_\kappa f$ then

(i) $0 \leq f < 1$.
(ii) $T_\kappa f \geq f$ with strict inequality when $f(x) > 0$.
(iii) $T_\kappa f \leq f/(1 - f)$ with strict inequality when $f(x) > 0$.
(iv) If $\kappa$ is irreducible then either $f = 0$ everywhere or $f > 0$ almost everywhere.

Proof. (i) $f(x) = 1 - e^{-T_\kappa f(x)} \leq 1$ so

$$(T_\kappa f)(x) \leq (T_\kappa 1)(x) \leq \int_S \kappa(x, y) d\mu(y) < \infty$$

by the assumption in (??) and it follows that $f(x) < 1$. (ii) This is immediate from $f = 1 - e^{-T_\kappa f} \leq T_\kappa f$ with equality only when $T_\kappa f = 0$. (iii) $e^{-T_\kappa f} = 1 - f$. Since $f < 1$, $e^{T_\kappa f} = 1/(1 - f)$ and we have

$$T_\kappa f \leq e^{T_\kappa f} - 1 = \frac{1}{1 - f} - 1 = \frac{f}{1 - f}$$

with equality only when $T_\kappa f = 0$.

(iv) Let $A = \{x \in S : f(x) = 0\}$. For $x \in A$, $(\Phi_\kappa f)(x) = f(x) = 0$. It follows that if $x \in A \kappa(x, y) = 0$ for $y \in S - A$, and the conclusion follows from the assumption of irreducibility.

Lemma 3.4.5. Suppose that $\kappa$ is irreducible. Suppose that $f = \Phi_\kappa f$ and $g = \Phi_\kappa g$ with $0 \leq f \leq g$ then either $f = 0$ or $f = g$.

Proof. By part (iv) of Lemma 3.4.4 we can suppose that $f > 0$ a.e. Let $h = (g - f)/2 \geq 0$ so that $f + h = (f + g)/2$. The function $t \to 1 - e^{-t}$ is strictly convex so

$$1 - e^{-(u+v)/2} \geq \frac{1}{2} (1 - e^{-u}) + \frac{1}{2} (1 - e^{-v})$$

and we have

$$\Phi_\kappa((f + g)/2) = 1 - e^{-T_\kappa(f+g)/2} \geq 1 - e^{-[T_\kappa(f) + T_\kappa(g)]/2}$$

$$\geq \frac{1}{2} (1 - e^{-T_\kappa f}) + \frac{1}{2} (1 - e^{-T_\kappa g}) = \frac{1}{2} (f + g) = f + h$$

where the $\leq$ is strict inequality at every point where $f < g$ and thus $\Phi_\kappa f < \Phi_\kappa g$ and $T_\kappa f < T_\kappa g$. On the other hand

$$1 - \Phi_\kappa((f + g)/2) = e^{-T_\kappa(f+h)} = e^{-T_\kappa f} e^{-T_\kappa h}$$

$$= (1 - f) e^{-T_\kappa h} = (1 - f)(1 - T_\kappa h)$$

Combining the last two computations

$$(1 - f)(1 - T_\kappa h) \leq 1 - (f + h) = 1 - f - h$$
and thus \((1 - f)T_\kappa h \geq h\) with strict inequality when \(g > f\).

Suppose now that \(g > f\) on a set of positive measure. Then the last inequality and (iii) of Lemma 3.4.4 imply

\[
\int_S fT_\kappa h \, d\mu > \int_S f \cdot \frac{h}{1 - f} = \int_S h \cdot \frac{h}{1 - f} \geq \int_S hT_\kappa f \, d\mu \tag{3.4.2} \]

The integrals are finite because \(\kappa \in L^1\) and \(f, h \leq 1\). However, since \(\kappa\) is symmetric \(T_\kappa\) is a symmetric operator contradicting (3.4.2) which implies \(f = g\) a.e. \(\square\)

Taking \(g = \rho_\kappa\) which is a solution

**Lemma 3.4.6.** If \(\kappa\) is irreducible then \(f = 0\) and \(f = \rho_\kappa\) are the only solutions. These solutions may coincide.

**Lemma 3.4.7.** If \(\|T_\kappa\| \leq 1\) then \(\rho_\kappa = 0\).

*Proof.* Suppose that \(f\) is a solution and that we do not have \(f = 0\) a.e. (ii) in Lemma 3.4.4 implies that \(T_\kappa f \geq f\) with \(T_\kappa > f\) on a set of positive measure so \(\|T_\kappa\| > 1\) contradicting the assumption \(\|T_\kappa\| \leq 1\) so \(f = 0\) a.e. and \(f = \Phi_\kappa f = 0\). \(\square\)

To approximately calculate \(S_\infty\) Riordan uses (see page 912)

**Fact 1** Lemma 3.4.8. If \(\phi : (0, 1] \to [0, 1]\) satisfies \(F(\phi) \geq \phi\) then \(S_\infty \geq \phi\).

**Fact 2** Lemma 3.4.9. If \(\phi : (0, 1] \to [0, 1]\) satisfies \(F(\phi) \leq \phi\) and \(F(\phi)\) is not \(\equiv 0\) then \(S_\infty \leq \phi\).
3.5 Survival probabilities for examples

3.5.1. Square-root model

If the initial type is $a$ then the expected number of first generation children that survive is

$$\mu = \int_0^1 \frac{c}{\sqrt{ab}} S_\infty(b) \, db$$

The set of offspring is a Poisson process on $(0, 1]$. We thin the process by keeping only the points that survive. Survival events are independent and independent of the Poisson process so the number of survivors has a Poisson distribution.

If the number that survive is Poisson with mean $\mu$ the probability at least one survives is $1 - e^{-\mu}$ so

$$S_\infty(a) = 1 - \exp \left( - \int_0^1 \frac{c}{\sqrt{ab}} S_\infty(b) \, db \right) \quad \text{(3.5.1)}$$

If we let $A_c = \int_0^1 (c/\sqrt{b}) S_\infty(b) \, db$ then

$$S_\infty(a) = 1 - \exp(-A_c/\sqrt{a})$$

so using (3.5.1), equating the exponentials, and multiplying by $\sqrt{a}$ we have

$$A_c = c \int_0^1 \frac{1}{\sqrt{b}} \left[ 1 - \exp(-A_c/\sqrt{b}) \right] \, db \quad \text{(3.5.2)}$$

We want to solve for $A_c$, or at least determine the behavior of $A_c$ as $c \to 0$. Let

$$g(x) = \int_0^1 \frac{1}{\sqrt{b}} [1 - \exp(-x/\sqrt{b})] \, db$$

With this notation $A_c = cg(A_c)$.

**Lemma 3.5.1.** $g(x) \sim 2x \log(1/x)$ as $x \to 0$.

**Proof.** The first step is to change variables $z = \sqrt{b}$, $dz = (1/2)b^{-1/2}db$ to get

$$g(x) = 2 \int_0^1 dz \left[ 1 - e^{-x/z} \right]$$

Noting that

$$1 - e^{-x/z} \approx \begin{cases} 
  x/z & z \gg x \\
  1 & z \ll x
\end{cases}$$

suggests an approach to approximating the integral

$$2 \int_0^{Kx} dz \left[ 1 - e^{-x/z} \right] \in [0, 2Kx]$$
If $K$ is large
\[
2 \int_{Kx}^{1} dz \left[ 1 - e^{-x/z} \right] \approx 2x \int_{Kx}^{1} \frac{dz}{z} = -2x \ln Kx = 2x \log(1/x) - 2x \log(K)
\]
Taking $K = \log \log(1/x)$ now gives the desired result.

To compute the asymptotics for the critical value now we note that
\[
c = \frac{A_c}{g(A_c)} \sim \frac{1}{2 \log(1/A_c)} \quad \text{or} \quad \log(1/A_c) \sim 1/2c
\]
which implies $A_c = \exp(-(1 + o(1))/2c)$

Riordan says one can get more precise results by using
\[
g(x) = 2 - 2e^{-x} + 2x \int_{x}^{1} \exp(-xt)/t \, dt
\]
This leads to the result given earlier in (3.2.10)
\[
f(c) \sim 2e^{1-\gamma} \exp(-1/2c)
\]

3.5.2. Chung-Lu model

Our first step is to repeat the argument above for a general rank-1 model
\[
\kappa(x, y) = c\psi(x)\psi(y).
\]
If the initial type is $a$ then the expected number of first generating children that survive is
\[
\mu = \int_{0}^{1} c\psi(a)\psi(b)S_{\infty}(b) \, db
\]
If the number that survive is Poisson with mean $\mu$ the probability at least one survives is
\[
1 - e^{-\mu}
\]
so
\[
S_{\infty}(a) = 1 - \exp\left(-\int_{0}^{1} c\psi(a)\psi(b)S_{\infty}(b) \, db\right) \quad (3.5.3) \quad \text{SaeqCL}
\]
If we let $A_c = \int_{0}^{1} c\psi(b)S_{\infty}(b) \, db$ then
\[
S_{\infty}(a) = 1 - \exp(-A_c\psi(a))
\]
so we have
\[
A_c = c \int_{0}^{1} \psi(b)[1 - \exp(-A_c\psi(b))] \, db \quad (3.5.4) \quad \text{AceqCL}
\]
Again if we let
\[
g(x) = \int_{0}^{1} \psi(b)[1 - \exp(-x\psi(b))] \, db \quad (3.5.5) \quad \text{gdefCL}
\]
With this notation $A_c = cg(A_c)$
3.5. SURVIVAL PROBABILITIES FOR EXAMPLES

$\gamma > 3$

In the Chung-Lu model $\psi(x) = x^{-1/(\gamma - 1)}$. When $\gamma > 3$

$$\int_0^1 \psi(x)^2 \, dx < \infty \quad (3.5.6)$$

so $c_c = 1/\int_0^1 \psi(x)^2 \, dx > 0$. If (3.5.6) holds than as $x \to 0$

$$g(x) \sim x \int_0^1 \psi(y)^2 \, dy$$

so, in contrast to the square root model, $g$ has a finite slope at 0.

$$g'(x) = \int_0^1 \psi^2(b) \exp(-x\psi(b)) \, d\beta$$

so $g'(x)$ decreases as $x$ increases i.e., $x$ is concave.

Since $g(0) = 0$, $g'(0) = \int_0^1 \psi(x)^2 \, dx$, and (3.5.5) implies that as $x \to \infty$

$$g(x) \to \int_0^1 \psi(b) \, db \quad \text{and hence} \quad g(x)/x \to 0.$$  

Thus we will have a solution of $x/c = g(x)$ only if $1/c < \int_0^1 \psi(x)^2 \, dx$, i.e., $c > c_c$. Expanding $g$ in power series about 0

$$g(x) = xg'(0) + \frac{x^2}{2}g''(0) + \cdots$$

so dropping the three dots (and leaving it to the reader to justify the computation that follows by replacing $g''(0)$ by $g''(y_x)$ with $y_x \in (0, x)$) we have

$$A_c = cg(A_c) \approx cg'(0)A_c + g''(0)\frac{A_c^2}{2}$$

Solving and recalling $c_c = 1/g'(0)$ we have

$$A_c = \frac{2cg'(0) - 1}{-g''(0)} = \frac{2g'(0)}{-g''(0)}(c - c_c)$$

If the constant looks weird recall that $g''(0) < 0$ since $g$ is concave.

$$S_\infty(a) = 1 - \exp(-A_c\psi(a)) \sim \psi(a)A_c$$

as $A_c \to 0$ so as $c \downarrow c_c$

$$S_\infty(a) \sim C\psi(a)(c - c_c)^\beta \quad \text{with} \ \beta = 1.$$  

So in the Chung-Lu model with $\gamma > 3$ the critical behavior of the survival probability has the same behavior in the Erdős-Rényi case.
2 < γ < 3

When γ = 3 we have ψ(x) = x^{-1/2} this essentially the square root model, so we move on to the case 2 < γ < 3 in which case

\[ g(x) = \int_0^1 \frac{1}{b^{1/(\gamma-1)}}(1 - \exp(-x/b^{1/(\gamma-1)}) \, db \]

Imitating the proof for the square-root model we change variables

\[ z = b^{1/(\gamma-1)}, \quad dz = \frac{1}{\gamma - 1} b^{1/(\gamma-1)-1} \]

which means

\[ db = (\gamma - 1)b^{-1/(\gamma-1)+1} \, dz = (\gamma - 1)z^{-1+(\gamma-1)} \, dz = (\gamma - 1) \frac{dz}{z^{2-\gamma}}. \]

Thus we have

\[ g(x) = (\gamma - 1) \int_0^1 \frac{1}{z^{3-\gamma}}(1 - \exp(-x/z)) \, dz \]

To deal with this formula we change variables again \( z = x/y, \, dz = -(x/y^2)dy \) to get

\[ g(x) = x^{\gamma-2} \int_0^\infty y^{1-\gamma}(1 - e^{-y}) \, dy \]

When γ > 2 we have 1 - γ < -1 so the integral is convergent at ∞. Near 0 we have \( 1 - e^{-y} \sim y \). The exponent \( 2 - \gamma < -1 \) when γ < 3 so the integral is convergent at 0 and the above is

\[ \sim C_\gamma x^{\gamma-2} \quad \text{where} \quad C_\gamma = \int_0^\infty y^{1-\gamma}(1 - e^{-y}) \, dy \]

As \( x \to 0, \, x/g(x) \sim x^{3-\gamma}/C_\gamma \). Using \( c = A_c/g(A_c) \) we have \( c \sim A^{3-\gamma}/C_\gamma \) and

\[ A_c \sim (C_\gamma x)^{1/(3-\gamma)} \]

Recall that \( A_c = \int_0^1 cb^{-1/(\gamma-1)}S_\infty(b) \, db \) so this is the survival probability starting from a random point chosen according to a measure that is finite when γ > 2.

Recall that in the configuration model the survival probability is \( c_k p^{(\gamma-2)/(3-\gamma)} \) when the degree distribution \( p_k \sim c k^{-\gamma} \). How are the two results related?
3.6 Component sizes in the subcritical case

3.6.1 Cases when the max is $O(\log n)$

Theorem 3.12 in BJR shows that if

$$(A) \quad \sup_{x,y} \kappa(x,y) < \infty$$

then in the subcritical regime the largest component is $O(\log n)$.

The proof given in Section 12 of their paper (see Theorem 12.5) is more than a little mysterious (to me at least) but if one is willing to assume instead

$$(A') \quad \text{There is a subcritical kernel } \bar{\kappa} \text{ for an inhomogeneous graph with finitely many types that has } \bar{\kappa}(x,y) \geq \kappa(x,y)$$

then the conclusion follows easily from the reasoning in Section 3.1.3.

Turova (2011)

For the rest of the section we will concentrate on the rank-1 case because in that situation we can define a cluster exploration process that is a random walk. For simplicity, we will work with the limiting object in which the children of $y$ in the branching process approximation are a Poisson process with mean measure $\kappa(y,z)\,dz$ on $(0,1]$. See Section 3.4 for a little more detail. In the rank-1 case $\kappa(y,z) = c\psi(y)\psi(z)$ and hence

$$c_c = 1/\int_0^1 \psi(y)^2\,dy$$

In order to have a subcritical phase we will suppose $\int \psi(y)^2\,dy < \infty$ Using the notation of Turova (2011), we let

$$M = \int \psi(u)\,du < \left(\int \psi(y)^2\,dy\right)^{1/2} < \infty$$

When $M < \infty$ the set of offspring of the vertex $y$ can be constructed in two steps.

(i) The total number of offspring is $N(y) = \text{Poisson with mean}$

$$\int \kappa(y,z)\,dz = c\psi(y)M.$$ 

(ii) Conditional on $N(y) = m$, the labels of the $m$ offspring are distributed according to $ar{\psi}(z) = \psi(z)/M$.

Given this procedure we can construct the exploration process as follows. If we start from $X_0 = x_0$ then the number of offspring $N(x_0) = \text{Poisson with mean } c\psi(x_0)M$. Since $A_0 = 1$ we have

$$A_1 = N(x_0)$$
CHAPTER 3. INHOMOGENEOUS RANDOM GRAPHS

For all $t \geq 1$ the vertex $X_t$ whose neighbors are added on the $t$th step is distributed according to $\tilde{\psi}(z)$, and

$$A_{t+1} = A_t - 1 + N(X_t)$$

The distribution of $N(X_t)$ is a mixture

$$\int_0^1 \tilde{\psi}(y) \text{Poisson}(c\psi(y)M) \, dy$$

At time $\tau = \inf\{t : A_t = 0\}$ we have found the $\tau$ members of the cluster.

The process is easier to analyze if it is homogeneous in time so we will suppose that $X_0$ is distributed according to $\bar{\psi}(z)$. To simplify typing let $Y_t = -1 + N(X_t)$. Recalling the formula for the Poisson and dropping the subscript $t$, the moment generating function of $Y$ is

$$\phi(\theta) \equiv E e^{\theta Y} = e^{-\theta} \int_0^1 \tilde{\psi}(y) e^{c\psi(y)M(e^\theta - 1)} \, dy$$

$W_t = \exp(\theta A_t)/\phi(\theta)^t$ is a martingale for $t \leq \tau$. To choose $\theta$ we note that

$$\phi'(\theta) = -\phi(\theta) + e^{-\theta} \int_0^1 \frac{\psi(y)}{M} e^{c\psi(y)M(e^\theta - 1)} \cdot c\psi(y)M e^\theta \, dy$$

Taking $\theta = 0$ we have

$$\phi'(0) = -1 + \int_0^1 c\psi(y)^2 dy < 0$$

since we are supposing $c < c_c$.

Since $\phi(0) = 1$ and $\phi'(0) < 0$, if we choose $\theta_0 > 0$ small enough then $\phi(\theta_0) < 1$. Using the optional stopping theorem for the nonnegative supermartingale $W_t$ at time $\tau$ which has $A_\tau = 0$

$$e^{\theta_0} \geq E(\phi(\theta_0)^{-\tau})$$

If we let $e^\alpha = 1/\phi(\theta_0) > 1$ then we have $E e^{\alpha \tau} \leq e^{\theta_0}$ and

$$P(\tau \geq k) \leq e^{\theta_0 e^{-\alpha k}}$$

The last result is for the cluster size when we start from a random vertex chosen according to $\tilde{\psi}(y)$. If we start from a fixed vertex $x$ then $A_1 = N(x)$ and $X_1$ has distribution $\bar{\psi}(y)$ so by (3.6.2)

$$P(\tau \geq k + 1|X_0 = x, N(x) = m) \leq e^{\theta_0 m} e^{-\alpha k}$$

Here the martingale starts when $A_1 = m$ instead of $A_0 = 1$. Multiplying by $P(N(x) = m)$ and summing over $m$

$$P(\tau \geq k + 1|X_0 = x) \leq e^{-\alpha k} E e^{\theta_0 N(x)} = e^{-\alpha k} \exp(c\psi(x)M(e^\theta - 1))$$

To bound this it is natural to assume that

$$(A1) \quad C_a = \int_0^1 \exp(a\psi(x)) \, dx < \infty \quad \text{for some } a > 0$$
3.6. COMPONENT SIZES IN THE SUBCRITICAL CASE

To be able to relate the integral to \( \sum_{i=1}^{n} \exp(a\psi(i/n)) \) we will assume that

\[ A2) \; x \to \psi(x) \text{ is decreasing.} \]

(Together these two assumptions are a little stronger than Turova’s Assumption 1.1.) If \( \theta_0 \) is small then \( c_x M(e^{\theta_0} - 1) \leq a \) so

\[
\frac{1}{n} \sum_{i=1}^{n} P(\tau \geq k + 1|X_0 = i/n) \leq e^{-a k} \int_{0}^{1} e^{a \psi(x)} \, dx
\]

so we have

\[
\sum_{i=1}^{n} P(\tau \geq k + 1|X_0 = i/n) \leq nC_a e^{-a k}.
\]

If we take \( k = ((1 + \epsilon)/\alpha) \log n \) then the probability that the largest cluster \( > (1 + \epsilon/\alpha) \log n \)

is \( \leq C_a / n^\epsilon \to 0 \) so

\[
P \left( \max_{1 \leq i \leq n} |C_i| > \frac{1 + \epsilon}{\alpha} \log n \right) \to 0 \quad (3.6.3) \text[Turbd2]
\]

Here we have gotten an upper bound on component sizes that are \( O(\log n) \). Turova (2011) obtained results that were asymptotically sharp

**Theorem 1.3. in Turova (2011).** Let \( \kappa(x, y) = c \psi(x) \psi(y) \) ad define \( r(c) \) by

\[
c^r = \left( \int \psi^2(x) \, dx \right)^{-1}
\]

As \( n \to \infty \) we have

\[
\frac{C_1(n, \kappa)}{\log n} \to \frac{1}{\log r(c)}
\]

Her Theorem 1.4 gives an explicit formula for \( r(c) \). Intuitively this is done by optimizing the \( \theta \) in the bound that comes from the moment generating function, but it takes ore than a little work to carry out the plan announced in the previous sentence. We refer the reader to her paper for details.
3.6.2. Cases where the max is a power of $n$

We will consider the Chung-Lu model introduced in Section 3.2.3. $\kappa(x, y) = c\psi(x)\psi(y)$ where $\psi(x) = x^{-a}$ with $a < 1/(\gamma - 1)$ and we assume $\gamma > 3$. To make it easier to do computations we will have self-loops at $i$ with probability $p(i, i)$ and assume they add only 1 to the degree of $i$ instead of 2.

A. Degree distribution

The probability of an edge between $i$ and $j$ is

$$p(i, j) = \kappa(i/n, j/n)/n = ci^{-a}j^{-a}n^{2a-1}$$

**Lemma 3.6.1.** As $n \to \infty$

$$E(d_i) \sim \frac{c}{1-a} \cdot \frac{n^a}{i^a}$$

**Proof.** The expected degree of $i$

$$E(d_i) = \frac{c}{1-a} \cdot \sum_{j=1}^{n} \frac{1}{j^a}$$

To evaluate the sum, we think about Riemann approximating sums for the integral of $1/x^a$ when the end points of the intervals are integers. Since $1/x^a$ is decreasing, if we evaluate at the left end point

$$\sum_{j=1}^{n} \frac{1}{j^a} \geq \int_{1}^{n+1} x^{-a} \, dx = \frac{(n+1)^{1-a} - 1}{1-a}$$

To get a bound in the other direction note that if we evaluate at the right endpoint

$$\sum_{j=2}^{n} \frac{1}{j^a} \leq \int_{1}^{n} x^{-a} \, dx \leq \frac{n^{1-a}}{1-a}$$

which proves the desired result.

Given the formula for the mean degree in Lemma 3.6.1 the next result should not be surprising, but to prove it we need estimates of the deviations of the degrees from their means.

**Lemma 3.6.2.**

$$P \left( d_1 = \max_{1 \leq i \leq n} d_i \right) \to 1$$ (3.6.4) \(\text{listmax}\)

$$d_1/Ed_1 \to 1 \text{ as } n \to \infty \text{ and hence}$$

$$\max_{1 \leq i \leq n} d_i / Ed_1 \to 1$$ (3.6.5) \(\text{maxdegree}\)
Proof. Our first step is to compute the variance.

\[
\text{var}(d_i) = \sum_{j=1}^{n} p(i, j)(1 - p(i, j)) \leq \sum_{i=1}^{n} p(i, j) = Ed_i
\]

Chebyshev’s inequality says that if \( E(Y^2) < \infty \)

\[
P(|Y - EY| > y) \leq \frac{\text{var}(Y)}{y^2}
\]  

(3.6.6) Cheby

so taking \( y = (Ed_i)^{2/3} \) we have

\[
P(|d_i - Ed_i| > (Ed_i)^{2/3}) \leq \frac{Ed_i}{(Ed_i)^{4/3}} \to 0.
\]

From this it follows that \( d_i/Ed_i \to 1 \) in probability, i.e., for any fixed \( i \) if \( \delta > 0 \)

\[
P(1 - \delta < d_i/Ed_i < 1 + \delta) \to 0
\]  

(3.6.7) LLNdi

as \( n \to \infty \). Since \( Ed_i = Ed_1/i^\alpha \), we can conclude that for any fixed \( I \)

\[
P\left(d_1 = \max_{1 \leq i \leq I} d_i\right) \to 1
\]  

(3.6.8) firsti

The argument cannot be used to prove (3.6.4) because the error bound from Chebyshev’s inequality is not good enough. To get a better bound we turn to the moment generating function. The analogue of Chebyshev’s inequality is:

**Lemma 3.6.3.** If \( \theta > 0 \) and the moment generating function, m.g.f., \( \phi(\theta) = Ee^{\theta Y} < \infty \) then

\[
e^{\theta y}P(Y \geq y) \leq \phi(\theta)
\]  

(3.6.9) expCh

Proof. This holds since \( e^{\theta y} \geq e^{\theta y} \) when \( Y \geq y \). 

Since oriented edges from \( i \) to \( j \) are independently present with probability \( p(i, j) \)

\[
Ee^{\theta d_i} = \prod_{j=1}^{n} \left(1 - p(i, j) + p(i, j)e^{\theta}\right) = \prod_{j=1}^{n} \left(1 + p(i, j)[e^{\theta} - 1]\right)
\]  

(3.6.10) edi

since the m.g.f. of a sum of independent random variables is the product of the m.g.f. We need to simplify the last expression. To do this we are guided by the fact that the sum of independent Bernoulli random variables is a approximately Poisson, while if \( Z \) is Poisson with mean \( \mu \) the mgf is

\[
E e^{\theta Z} = \sum_{k=0}^{\infty} e^{-\mu} \frac{\mu^k}{k!} e^{\theta k} = \exp(\mu[e^{\theta} - 1])
\]
Using $1 + x \leq e^x$ in (3.6.10) and letting $\mu_i = \sum_j p(i, j)$ we have

$$E e^{\theta d_i} \leq \exp(\mu_i[e^{\theta} - 1])$$

so using (3.6.9), if $c > 1$

$$P(d_i \geq c\mu_i) \leq \exp(\mu_i[1 - c\log 2])$$

This is a very powerful inequality but we have to manipulate it to be useful. A more systematic approach is to optimize over $\theta$ but that can get messy. To simplify our inequality we set $\theta = \log 2$ to get

$$P(d_i \geq \mu_1/2) \leq \exp\left(\mu_i \left[1 - \frac{\mu_1}{2\mu_i} \log 2\right]\right)$$

$\mu_1/\mu_i = i^a$. If $(\mu_1/2\mu_i) \log 2 > 2$ then for $i > I$ we have $(\mu_1/4\mu_i) \log 2 > 1$ and hence

$$P(d_i \geq \mu_1/2) \leq \exp\left(\mu_i \left[-\frac{\mu_1}{4\mu_i} \log 2\right]\right) = \exp\left(-\frac{\mu_1}{4} \log 2\right)$$

It follows that

$$\sum_{i=I+1}^{n} P(d_i \geq \mu_1) \leq n \exp\left(-\frac{\mu_1}{2} \log 4\right)$$

as $n \to \infty$ since $\mu_1 \sim Cn^{1-a}$. Combining this with (3.6.8) proves (3.6.4). (3.6.5) now follows from (3.6.7).

\[ \square \]

**B. Subcritical cluster sizes**

As in Section 3.6.1, we use the exploration process. If we start from $X_0 = x_0$ then the number of neighbors $N(x_0) = \text{Poisson with mean } c\psi(x_0)M$ where where $\psi(x) = x^{-1/(\gamma-1)}$, $M = \int_0^1 \psi(z) \, dz$. Since initially the active set $A_0 = 1$ we have

$$A_1 = N(x_0)$$

For all $t \geq 1$ the index of the vertex $X_t$ whose neighbors are revealed on the $t$th step is distributed according to $\tilde{\psi}(z) = psi(z)/M$, and

$$A_{t+1} = A_t - 1 + N(X_t). \quad (3.6.11)$$

The distribution of $N(X_t)$ is a mixture

$$\int_0^1 \tilde{\psi}(y) \text{Poisson}(c\psi(y)M) \, dy$$
3.6. COMPONENT SIZES IN THE SUBCRITICAL CASE

while the expected value is

\[ \mathbb{E}(X_t) = c \int_0^1 \psi(y)^2 dy < 1 \quad \text{if } c < c_c. \]  

(3.6.12) \[ \text{meanNX} \]

Using the reasoning from Section 3.2.3, when we compute the degree distribution we can think of \( \text{Poisson}(c\psi(y)M) \) as being a point mass at the mean. \( c\psi(y)M \geq x \) when

\[ y^{-1/(\gamma-1)} \geq x/cM \quad \text{or} \quad y \leq (x/cM)^{-(\gamma-1)} \]

Letting \( z = (x/cM)^{-(\gamma-1)} \) we have

\[ P(N(X_t) \geq x) = \int_0^z \frac{\psi(y)}{M} dy = \frac{C}{M} x^{1-1/(\gamma-1)} = C_5 x^{-(\gamma-1)+1} \]  

(3.6.13) \[ \text{xiidist} \]

where the two \( C \)'s are constants whose values are not important. The sudden appearance of the subscript 5 is to make it easier to compare with Section 2.5. Taking the derivative with respect to \( y \) we see that, as in the configuration model, the distribution of the number of vertices added is \( Cy \) times the original degree distribution.

Our next goal is to show that in the subcritical regime of the Chung-Lu model with \( \gamma > 3 \) the size of the largest cluster is of the same order as the largest degree.

**Theorem 3.6.4.** In the Chung-Lu model with \( \gamma > 3 \) then for any \( c < c_c \) there is a constant \( C_{\gamma,c} \) so that

\[ P\left( \max_v |C(v)| \geq C_{\gamma,c} n^{-1/(\gamma-1)} \right) \to 0 \]

By analogy with Janson’s result for the configuration model, (2.5.3), it is natural to expect that

\[ C_{\gamma,c} = Ed_1/(1 - \nu) \]

where \( \nu = \mathbb{E}(X_t) \).

**Proof.** We proved a version of Theorem 3.6.4 for the configuration model in Section 2.5. To be able to more easily compare with the previous proof we rewrite the recursion (3.6.11) to look like the one in Section 2.5

\[ S_{t+1} = S_t - 1 + \xi_t \]

where \( S_t = A_t \) and \( \xi_t = N(X_t) \). The new proof is almost the same as before but easier. We do not need the first part of the argument where we develop a comparison of the \( \xi_i \) with independent random variables \( X_i \).

If \( |C(x_0)| \geq M \) then \( \tau \geq M \) and

\[ S_M = d(x_0) + \sum_{i=1}^M (\xi_i - 1) \]  

(3.6.14) \[ \text{newSM} \]
which is (2.5.4). As noted in (3.6.12) we have $E\xi_i = \nu < 1$.

For some $A$ large and $\delta < 1/(\gamma - 1)$ small, to be chosen later, let

$$M = An^{1/(\gamma - 1)} \quad \text{and} \quad M_1 = n^{1/(\gamma - 1) - \delta}$$  \hspace{1cm} (3.6.15) \text{ Mdefs}

Note that $M$ is the desired upper bound on $C_1$. If we define the truncated random variables $Y_i = \xi_i 1(\xi_i \leq M_1)$ then $EY_i \leq E\xi_i = \nu$. If $|C(x_0)| > M$ then (3.6.14) holds and so using $\xi_i \leq \Delta = \max_i d_i$ and computing as in (2.5.12)

$$M \leq d(x_0) + \sum_{i=1}^{M} \xi_i \leq \Delta + \sum_{i=1}^{M} \xi_i 1(\xi_i \leq M_1) + \sum_{i=1}^{M} \xi_i 1(\xi_i > M_1)$$  \hspace{1cm} (3.6.16) \text{ newJ2.9}

$$\leq \sum_{i=1}^{M} Y_i + \Delta \left( 1 + \sum_{i=1}^{M} 1(\xi_i > M_1) \right)$$

$$\leq \nu M + \sum_{i=1}^{M} (Y_i - EY_i) + \Delta \left( 1 + \sum_{i=1}^{M} 1(\xi_i > M_1) \right)$$

Thus if $\nu < 1 - 2\varepsilon$ (which holds for $\varepsilon = (1 - \nu)/3$)

$$P(|C(x_0)| \geq M) \leq P \left( \sum_{i=1}^{M} |Y_i - EY_i| > \varepsilon M \right) + P \left( \sum_{i=1}^{M} 1(\xi_i > M_1) > \frac{\varepsilon M}{\Delta} - 1 \right)$$  \hspace{1cm} (3.6.17) \text{ newJ2.10}

For the first sum fix a number $r \geq \gamma$ so that $r\delta > 2$. Rosenthal’s inequality for nonnegative random variables given in (2.5.14) states

$$E \left( \sum_{k=1}^{n} Z_i \right)^r \leq C_r \max \left( \left( \sum_{k=1}^{n} EZ_i^2 \right)^{r/2}, \sum_{k=1}^{n} EZ_i^r \right)$$

Letting $Z_i = |Y_i - EY_i|$ we have

$$E \left( \sum_{k=1}^{n} |Y_i - EY_i| \right)^r \leq C_r \left( (nE|Y_i - EY_i|^2)^{r/2} + nE|Y_i - EY_i|^r \right)$$

Taking $n = M$ we have

$$E \left| \sum_{i=1}^{M} |Y_i - EY_i| \right|^r \leq C_6 M^{r/2} (E|Y_1 - EY_1|^2)^{r/2} + C_7 M E|Y_1 - EY_1|^r$$

$$\leq C_6 M^{r/2} (EY_1^2)^{r/2} + C_8 M EY_1^r$$  \hspace{1cm} (3.6.18) \text{ newJ2.11}

To go from the first line to the second we note that $E(X - c)^2$ is minimized when $c = EX$. For the second term writing $E(Y'; A)$ for the integral of $Y$ over $A$

$$E|X - EX|^r = E(|X - EX|^r; X \geq EX) + E(|X - EX|^r; X < EX) \leq E(|X|^r; X \geq EX) + |EX|^r P(X < EX) \leq 2E|X|^r.$$
where on the second line we have used the fact that $X \geq 0$ and then Jensen’s inequality to put the $r$th power inside the expected value.

We estimate the second moment of $Y_1$ using (3.6.13)

$$EY_1^2 = \int_0^\infty 2xP(Y_1 > x)\,dx = \int_0^{M_1} 2xP(\xi_1 > x)\,dx$$

$$\leq 1 + C_5 \int_1^{M_1} 2x^{3-\gamma}\,dx \leq C_9 M_1$$  \hspace{1cm} (3.6.19)  \hspace{1cm} \text{newJ2.12}

since $\gamma > 3$. Trivially $EY_1^r \leq M_1^r$. Using Markov’s inequality, (3.6.18), the second moment estimate in (3.6.19), $Y_1 \leq M_1$ and $M_1/M = A^{-1}n^{-\delta}$ we get (here $C_{10} = c_6\epsilon^{-r}C_6$ and $C_{11} = c_6\epsilon^{-r}C_8$ but $\epsilon$ is fixed)

$$P\left(\sum_{i=1}^M |Y_i - EY_i| > \epsilon M\right) \leq (\epsilon M)^{-r}E\left[\sum_{i=1}^M |Y_i - EY_i|\right]^r$$

$$\leq C_{10}(EY_1^2/M)^{r/2} + C_{11}M^{1-r}EY_1^r$$  \hspace{1cm} (3.6.20)  \hspace{1cm} \text{newJ2.14}

$$\leq C_{12}(M_1/M)^{r/2} + C_{13}M(M_1/M)^r$$

$$\leq C_{14}n^{-r\delta/2} + C_{15}n^{1-r\delta} = o(n^{-1})$$

since $r\delta > 2$.

For the second sum in (3.6.17) write $I_i = 1_{(\xi_i > M)}$ and following (2.5.17) conclude that

$$P\left(\sum_{i=1}^M I_i \geq L\right) \leq \binom{M}{L}P(I_k = 1, 1 \leq k \leq L)$$

$$= \binom{M}{L}P(I_1 = 1)^L = \binom{M}{L}P(X > \xi_1)^L \leq (MP(\xi_1 > M_1))^L$$  \hspace{1cm} (3.6.21)  \hspace{1cm} \text{newJ2.15}

Now by the tail bound on $\xi_i$ in (3.6.13) and the choices of $M$ and $M_1$ in (3.6.15)

$$MP(\xi_1 > M_1) \leq MC_5M_1^{-(\gamma - 1) + 1}$$

$$\leq C_5A n^{1/(\gamma - 1) + (2-\gamma)[1/(\gamma - 1)-\delta]} = C n^{(\gamma - 2)\delta-(\gamma - 3)/(\gamma - 1)}$$

Choose $\delta > 0$ so that $\delta_1 = (\gamma - 3)/(\gamma - 1) - (\gamma - 2)\delta > 0$. Then $MP(\xi_1 > M_1) = O(n^{-\delta_1})$ and

$$P\left(\sum_{i=1}^M I_i \geq L\right) = O(n^{-L\delta_1})$$  \hspace{1cm} (3.6.22)  \hspace{1cm} \text{newJ2.16}

If we choose $L > 1/\delta_1$ so that the last error term is $o(n^{-1})$. (3.6.5) and Lemma 3.6.1 imply that

$$\Delta/Ed_1 \to 1 \quad Ed_1 \sim cN^a/(1-a)$$
where $a = 1/\left(\gamma - 1\right)$. By (3.6.15) $M = An^{1/(\gamma - 1)}$ so if $A$ is chosen large enough

$$
\varepsilon \frac{M}{\Delta} - 1 > L \quad \text{for large } n. \tag{3.6.23}
$$

Combining this with (3.6.22) and (3.6.20) we have shown that the right-hand side of (3.6.17) is $o(n^{-1})$, which gives the conclusion of Theorem 3.6.4.

To show that all of our constants can be chosen the way we want them, we note that $\varepsilon = (1 - \nu)/3$ where $\nu = E\xi_i$, $\delta$ was chosen small enough so that

$$
\frac{\gamma - 3}{\gamma - 1} - \delta(\gamma - 2) > 0
$$

$r \geq \gamma$ was chosen so that $r\delta \geq 2$, and finally $A$ was chosen to give (3.6.23).
References


