UNC Lecture 2

General degree distributions

Rick Durrett
Motivation

In an Erdös-Rényi random graph, vertices have degrees that have asymptotically a Poisson distribution. However, in social and communication networks, the distribution of degrees is much different from the Poisson and in many cases has a power law form,

$$p_k \sim Ck^{-\gamma} \text{ as } k \to \infty.$$

At the turn of this century attention focused on a number of examples.

By the world wide web, we mean the collection of web pages and the oriented links between them. Barabási and Albert (1999) found that the in-degree and out-degrees have power laws with $\gamma_{\text{in}} = 2.1$, $\gamma_{\text{out}} = 2.7$.

By the Internet, we mean the physically connected network of routers that move email and files around the Internet. In 2000 the degree distribution that could be fit by a power law with $\gamma = 2.3$. 
The **movie actor network** in which two actors are connected by an edge if they have appeared in a film together has a power law degree distribution with $\gamma = 2.3$.

The **collaboration graph** in a subject is a graph with an edge connecting two people if they have written a paper together. Barabási et al. (2002) studied papers in mathematics and neuroscience published in 1991–1998. The fitted power laws had $\gamma_M = 2.4$ and $\gamma_{NS} = 2.1$.

**sex in Sweden** Liljeros et al. (2001) gathered data in a study of sexual behavior of 4,781 Swedes, and found that the number of partners per year had $\gamma_{male} = 3.3$ and $\gamma_{female} = 3.5$.

Note that in the first four examples $2 < \gamma < 3$
Configuration model

Let $d_1, \ldots, d_n$ be independent and have $P(d_i = k) = p_k$. Since we want $d_i$ to be the degree of vertex $i$, we condition on $E_n = \{d_1 + \cdots + d_n \text{ is even}\}$. To build the graph we think of $d_i$ half-edges attached to $i$ and then pair the half-edges at random. The picture gives an example with 8 vertices.
This construction can produce multiple edges and even self-loops but if \( p_k \) has a finite second moment the number is \( O(1) \)

**Theorem 2.1.2.** Let \( \mu = \sum_k k p_k \) and \( \mu_2 = \sum_k k(k - 1)p_k \). As \( n \to \infty \), the number of self-loops \( \chi_0 \) and the number of parallel edges \( \chi_1 \) are asymptotically independent Poisson\( (\mu_2/2\mu) \) and Poisson\( ((\mu_2/2\mu)^2) \).

In particular the probability of \( H_0 = \) “the resulting graph has no self-loops or parallel edges” converges to a positive limit as \( n \to \infty \).
Size-biased distribution

If we start with a given vertex $x$ then the number of neighbors (the first generation in the branching process) has distribution $p_j$. A first generation vertex with degree $k$ is $k$ times as likely to be chosen as one with degree 1, so the distribution of the number of children of a first generation vertex is for $k \geq 1$

$$q_{k-1} = \frac{kp_k}{\mu} \text{ where } \mu = \sum_k kp_k.$$  Let $\nu = \sum_k kq - k$

Sanity check. If $p_k = e^{-\lambda} \frac{\lambda^k}{k!}$, then $\mu = \lambda$, so

$$q_{k-1} = e^{-\lambda} \frac{k \lambda^k}{k! \lambda} = e^{-\lambda} \frac{\lambda^{k-1}}{(k-1)!} = \text{Poisson}(\lambda)$$
Phase transition

For $m \geq 1$, $EZ_m = \mu \nu^{m-1}$ so when $\nu < 1$

$$E \left( \sum_{m=0}^{\infty} Z_m \right) = 1 + \sum_{m=1}^{\infty} \mu \nu^{m-1} = 1 + \frac{\mu}{1 - \nu}$$

Let $G_0(z) = \sum_k p_k z^k$ and $G_1(z) = \sum_k q_k z^k$. Probability that the family of a first generation particle dies out satisfies $\rho_1 = G_1(\rho_1)$, and the probability the two phase process dies out is $\sum_{k=0}^{\infty} p_k \rho_1^k = G_0(\rho_1)$

**Theorem 2.1.3.** The condition for the existence of a giant component is $\nu > 1$. In this case the fraction of vertices in the giant component is asymptotically $1 - G_0(\rho_1)$. 
A more general approach

Molloy and Reed (1985), Janson and Luczak (2000).
For each $n$ we have a degree sequence $d_i^n$

(i) $n_k/n = |\{i : d_i = k\}|/n \to p_k$ as $n \to \infty$
(ii) $\lambda = \sum_k kp_k \in (0, \infty)$
(iii) $\sum_i d_i^2 = O(n)$ (implies $\sum_k kn_k/n \to \lambda$)
(iv) $p_1 > 0$

Why do this? Configuration model (CM) is fragile this definition is robust.
If we take a CM with $p_k = \text{Poisson} \left( \right)\lambda\text{,}$ result is not Erdös-Rényi.
If we take a CM and keep edges with probability $r$ the result is not a CM.
$d$ independent perfect matchings gives a non CM $d$-regular graph.
Subcritical cluster sizes


$$\mu_n = ED_n = \frac{1}{n} \sum_{k=0}^{\infty} kn_k = \frac{1}{n} \sum_{i=1}^{n} d_i^n$$

$$\nu_n = \frac{1}{n \mu_n} \sum_{k=0}^{\infty} k(k-1)n_k = \frac{1}{n \mu_n} \sum_{i=1}^{n} d_i^n (d_i^n - 1)$$

**Theorem 2.3.1.** Suppose $\mu_n \to \mu_{\infty} > 0$, $\nu_n \to \nu_{\infty} < 1$, and

$$P(D_n \geq k) \leq c_1 k^{1-\gamma}$$

for some $\gamma > 3$ and $c_1 < \infty$, then there is a constant $A$ so that

$$|C^{(1)}| \leq An^{1/(\gamma-1)}$$
Theorem 2.3.2. Let $\Delta_n = \max_{1 \leq i \leq n} d_i^n$ be the maximum degree

$$|C^{(1)}| = \frac{\Delta_n}{1 - \nu_n} + o_p(n^{1/(\gamma-1)})$$

To explain this result note that if $x$ is the vertex with largest degree and $\Delta_n = m$ then the law of large numbers implies that the number of vertices at distance 2 from $x$ in $C_x$ is $\sim mn\nu_n$, or more generally the number of vertices at distance $k$ from $x$ is $\sim mk^{-1}\nu_n$.

If $P(D_n \geq k) \leq c_1 k^{1-\gamma}$ then

$$P(\Delta_n \geq An^{1/(\gamma-1)}) \leq n \cdot c_1 A^{1-\gamma} n^{-1}$$
Exploration process

Fix a vertex $v$ and explore the component $C(v)$. We construct the random configuration as we need it. The process begins by declaring $v$ used and all the half-edges at $v$ to be active.

Then for $i = 1, 2, \ldots$, as long as there is an active half-edge, we pick one at random $x_i$. We choose its partner half-edge $y_i$ uniformly among all the half-edges except ($x_i$ and the ones that have been paired) and let $v_i$ be its vertex. If the vertex is not already used we declare its other $d(v_i) - 1$ half-edges active. Finally declare $x_i$, $y_i$, $v_i$ used. Repeat.

$S_0 = d(v)$. $S_i = S_{i-1} - 1 + \xi_i$, $\xi_i$ has the size biased distribution
Random walk estimates

\[ S_0 = d(v). \quad S_i = S_{i-1} - 1 + \xi_i, \]

\( \xi_i \) size biased distribution

\( \xi_i \) not independent but \( \leq X_i \) independent, \( P(X_i \geq x) \leq Cx^{2-\gamma} \)

Having passed to independent \( X_i \) it is now a question of random walk estimates but we only have \( 2 + \delta \) moments for \( X_i \geq 0 \), so we have to truncate.
van der Hofstad, Hooghiemstra, and Van Mieghem (2004)

$Z_m$ be the two phase branching process, $Z_m/(\mu\nu^{m-1} \rightarrow W$

**Theorem 2.3.1.** Suppose $\nu > 1$. Let $H_n$ be the distance between two randomly chosen points $x$ and $y$ on the giant component. For $k \geq 1$, let $a(k) = [\log_\nu k] - \log_\nu \in (-1, 0]$. As $n \to \infty$

$$P(H_n - [\log_\nu n] = k|H_n < \infty) = P(R_{a(n)} = k) + o(1)$$

If $\kappa = \mu(\nu - 1)^{-1}$ then for $a \in (-1, 0]$ 

$$P(R_a > k) = E(\exp(-\kappa\nu^{a+k}W_1W_2)|W_1W_2 > 0)$$

where $W_1$ and $W_2$ are independent copies of $W$.

Proof start bp at $x$ and $y$. Note fluctuations are $O(1)$
van der Hofstad, Hooghiemstra, and Znamenki (2004). Let $H_n$ be the distance between two randomly chosen points $x$ and $y$ on the giant component. If $\alpha = \gamma - 2 \in (0, 1)$ then

$$H_n \sim 2(\log \log n)/(-\log \alpha)$$

**Theorem 2.4.1.** Davies (1978). Consider a branching process with offspring distribution $\xi$ with $P(\xi > k) \sim B_\alpha k^{-\alpha}$ where $\alpha = \beta - 2 \in (0, 1)$. As $m \to \infty$, $\alpha^m \log(Z_m + 1) \to W$ with $P(W = 0) = \rho$ the extinction probability for the branching process.

$Z_m = n$ when $(\log n)/W = \alpha^{-m}$. The 2 comes from the fact that we grow from $x$ and from $y$. 
Percolation

Percolation is one of the simplest processes to study on random graphs. If we take a random graph and delete edges with probability $1 - p$ then we have another random graph.

Suppose that the original degree distribution was $p_k$. If we keep edges with probability $p$ then the probability an individual chosen at random will have $j$ neighbors after thinning is

$$
\hat{p}_j = \sum_{k=j}^{\infty} p_k \binom{k}{j} p^j (1 - p)^{(k-j)}
$$

and the mean degree is $\hat{\mu} = p \mu$.

Let $\hat{G}_0$ be the generating function of $\hat{p}_k$

$$
\hat{G}_0(z) = G_0(pz + (1 - p))
$$
As we explore the cluster containing a vertex, individuals in the first and subsequent generations will have $k$ neighbors (excluding their parent) with probability $q_k = (k + 1)p_{k+1}/\mu$ for $k \geq 0$. The probability that they will have $j$ children after thinning is

$$\hat{q}_j = \sum_{k=j}^{\infty} q_k \binom{k}{j} p^j (1 - p)^{k-j}$$

and the mean degree is $\hat{\nu} = p\nu$. So $p\nu > 1$ is the condition for a giant component.

Let $\hat{G}_1$ be the generating function of $\hat{q}_k$

$$\hat{G}_1(z) = G_1(pz + (1 - p))$$
## Percolation probability

Size of giant component be computed using generating functions as before. Let $\theta = p \nu$

<table>
<thead>
<tr>
<th>finite variance</th>
<th>$p_c$</th>
<th>perc. prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0 &lt; \gamma &lt; 4$</td>
<td>$&gt; 0$</td>
<td>$\sim C(\theta - 1)$</td>
</tr>
<tr>
<td>$\gamma = 3$</td>
<td>$= 0$</td>
<td>$\exp(- (1 + o(1)/cp))$</td>
</tr>
<tr>
<td>$2 &lt; \gamma &lt; 3$</td>
<td>$= 0$</td>
<td>$\sim cp^{(\gamma - 2)/(3 - \gamma)}$</td>
</tr>
</tbody>
</table>


Internet is robust to random attacks ($p_c = 0$) yet fragile, since it is easily destroyed by targeted attacks (take out highest degree nodes).

Percolation probabilities show that the part of the internet which survives can be very small. Bollobás and Riordan (2004).
Critical exponents vs. $\gamma$
Exercise 9 on page 425 of the third edition of Feller volume I. (edited)

Plane random walk with reflecting barriers. Consider a symmetric random walk on a subgraph of the two-dimensional integer lattice. The boundary is reflecting in the sense that, whenever the particle in the unrestricted random walk the particle would leave the region, it is not allowed to jump. Show that if every point in the region can be reached from every other point, and the region has \( K < \infty \) points there is a stationary distribution with \( u(x) = 1/K \), \( x \in D \). (If \( D \) is unbounded then the states are persistent null states and \( u(x) = 1 \), \( x \in D \) is a stationary distribution.)

Challenge. Prove the claim in parentheses.