Generating Functions and Degree Distributions

For degree and excess degree distributions we define generating functions

\[ g_0(z) = \sum_{k=0}^{\infty} p_k z^k \quad \text{and} \quad g_1(z) = \sum_{k=0}^{\infty} q_k z^k, \]

respectively.

They are not independent

\[ g_1(z) = \frac{1}{\langle k \rangle} \sum_{k=0}^{\infty} (k + 1) p_{k+1} z^k = \frac{1}{\langle k \rangle} \sum_{k=0}^{\infty} k p_k z^{k-1} = \frac{1}{\langle k \rangle} \frac{d g_0}{d z} = \frac{g'_0(z)}{g'_0(1)} \]

we add zero term because of infinity

Example (Poisson): \( p_k = e^{-c} \frac{c^k}{k!} \) \( \Rightarrow \) \( g_0(z) = e^{c(z-1)} \) and \( g_1(z) = e^{c(z-1)} \)

Example (power-law): \( p_k = C k^{-\alpha} \) \( \Rightarrow \) \( g_0(z) = \frac{Li_{\alpha-1}(z)}{\zeta(\alpha)} \). Thus,

\[ g_1(z) = \frac{Li_{\alpha-1}(z)}{z Li_{\alpha-1}(1)} = \frac{Li_{\alpha-1}(z)}{z \zeta(\alpha - 1)} \]
**Number of second neighbors of a vertex**

Probability that $i$ has exactly $k$ second neighbors

\[ p_k^{(2)} = \sum_{m=0}^{\infty} p_m P^{(2)}(k|m) \]

- **Degree distribution**: \( P^{(2)}(k|m) = \sum_{j_1=0}^{\infty} \cdots \sum_{j_m=0}^{\infty} \delta \left( k, \sum_{r=1}^{m} j_r \right) \prod_{r=1}^{m} q_{j_r} \)

  - All sets of values $j_1, j_2, ..., j_m$ that sum to $k$

- **Generating function**: \( g^{(2)}(z) = \sum_{k=0}^{\infty} p_k^{(2)} z^k = \sum_{k=0}^{\infty} z^k \cdot \sum_{m=0}^{\infty} p_m \sum_{j_1=0}^{\infty} \cdots \sum_{j_m=0}^{\infty} \delta \left( k, \sum_{r=1}^{m} j_r \right) \prod_{r=1}^{m} q_{j_r} = \cdots = \sum_{m=0}^{\infty} p_m \cdot \left( \sum_{j=0}^{\infty} q_j z^j \right)^m = g_0 \left( g_{1}(z) \right) \)
**Conclusion:** Once we know generating functions of $g_0$ and $g_1$ the generating function of second neighbor distribution is straightforward to calculate. Moreover, this can be extended to

$$g^{(3)}(z) = \sum_{k=0}^{\infty} \sum_{m=0}^{\infty} p_m^{(2)} P^{(3)}(k|m) z^k = \sum_{n=0}^{\infty} p_m^{(2)} (g_1(z))^m = g^{(2)}(g_1(z)) = g_0(g_1(g_1(z)))$$

$$\implies g^{(d)}(z) = g^{(d-1)}(g_1(z)) = g_0(\cdots g_1(z) \cdots)$$

**Problem:** Sometimes it is difficult to extract explicit probabilities for numbers of second neighbors and it is hard to evaluate $n$ derivatives (in order to recover the probabilities).

**Solution:** calculate the average number of neighbors at distance $d$. At $z=1$ of the first derivative we can evaluate the average of a distribution (see Slide 16).

$$\frac{dg^{(2)}}{dz} = g_0'(g_1(z))g_1'(z) \quad z=1, g_1(1)=1 \quad \implies c_2 = g_0'(1)g_1'(1) \quad g_0'(1)=\langle k \rangle \quad g_1'(k) = \sum_{k=0}^{\infty} kq_k =$$

\[
\frac{1}{\langle k \rangle} \sum_{k=0}^{\infty} k(k+1)p_{k+1} = \frac{1}{\langle k \rangle} (\langle k^2 \rangle - \langle k \rangle)
\]

**Conclusion:** $c_2 = \langle k^2 \rangle - \langle k \rangle$ and more general

$$c_d = \left( \frac{c_2}{c_1} \right)^{d-1} \quad c_1 \implies \quad \text{Condition of giant component's existance in configuration model is } \langle k^2 \rangle - 2\langle k \rangle > 0$$

[MR] A critical point for random graphs with given degree sequence
Let's use theory for more practical results ...

Given a network with \textbf{power-law distribution} \( p_k = Ck^{-\alpha}, \; \alpha > 0, \; k > 0 \)
Reminder: \( C \) is calculated from normalization condition, i.e., \( C = 1/\zeta(\alpha) \)

\[
p_k = \begin{cases} 
0 & k = 0 \\
k^{-\alpha}/\zeta(\alpha) & k > 0 
\end{cases}
\]

This network will have a giant component iff \( \langle k^2 \rangle - 2\langle k \rangle > 0 \)

\[
\langle k \rangle = \sum_{k=0}^{\infty} kp_k = \frac{1}{\zeta(\alpha)} \sum_{k=1}^{\infty} k^{-\alpha+1} = \frac{\zeta(\alpha-1)}{\zeta(\alpha)}
\]

\[
\langle k^2 \rangle = \sum_{k=0}^{\infty} k^2 p_k = \frac{1}{\zeta(\alpha)} \sum_{k=1}^{\infty} k^{-\alpha+2} = \frac{\zeta(\alpha-1)}{\zeta(\alpha)}
\]

\[\implies \zeta(\alpha - 2) > 2\zeta(\alpha - 1)\]
Figure 13.8: Graphical solution of Eq. (13.138). The configuration model with a pure power-law degree distribution (Eq. (13.133)) has a giant component if $\zeta(\alpha - 2) > 2\zeta(\alpha - 1)$. This happens for values of $\alpha$ below the crossing point of the two curves.
Happy families are all alike, every unhappy family is unhappy in its own way.

Leo Tolstoy

Fundamental theoretical and practical questions
- What are the fundamental processes that form a network?
- How to predict its future structure?
- Why should network have property X?
- Will my algorithm/heuristic work on networks created by similar processes?

Is it similar to the original network?
Rich-get-richer effect

Herbert Simon
1916-2001

Analyzed the power laws in economic data, suggested explanation of wealth distribution: return of investment is proportional to the amount invested, i.e., wealthy people will get more and more

Simon (1976). “On a class of skew distribution functions"

Derek Price
1922-1983

Studied information science; in particular, citation networks; his main assumption was about newly appearing papers that cite old papers with probability proportional to the number of citations those old papers have → the model is similar to Simon’s model.

Price (1976). "A general theory of bibliometric and other cumulative advantage processes"
Price’s model

$c$ on the average

New
Consider adding a single vertex $v$ in Price’s model, where $p_q(n)$ is the fraction of vertices in the network with in-degree $q$.

Probability of $v \rightarrow i$ citation is

$$\frac{q_i + a}{\sum_i (q_i + a)} = \frac{q_i + a}{n\langle q \rangle + na} = \frac{q_i + a}{n(c + a)}$$

Expected number of new citations to all nodes with degree $q$ is

$$np_q(n) \cdot c \cdot \frac{q + a}{n(c + a)} = \frac{c(q + a)}{c + a} p_q(n)$$

Thus, the number of vertices with in-deg $q$ after adding $v$ is

$$(n + 1) p_q(n + 1) = n p_q(n) + \frac{c(q - 1 + a)}{c + a} p_{q-1}(n) - \frac{c(q + a)}{c + a} p_q(n)$$

$$\implies p_q = \frac{q + a - 1}{q + a + 1 + a/c} p_{q-1} \implies p_q (q + a)^{-\alpha}$$

Use properties of gamma and beta functions.
Figure 14.2: Degree distribution in Price’s model of a growing network. (a) A histogram of the in-degree distribution for a computer-generated network with $c = 3$ and $a = 1.5$ which was grown until it had $n = 10^8$ vertices. The simulation took about 80 seconds on the author’s computer using the fast algorithm described in the text. (b) The cumulative distribution function for the same network. The points are the results from the simulation and the solid line is the analytic solution, Eq. (14.34).
Preferential Attachment (Barabasi-Albert)

- Initialize network with $m_0$ nodes ($m_0 \geq 2$, $d(i) \geq 1$)
- Add node $i$, connect it to exactly $c$ out of $m$ existing nodes with probability
  \[
  \Pr[i \rightarrow j] = \frac{k_j}{\sum_l k_l}
  \]
- Repeat previous step or stop if $|V| = n$

$p_k \sim k^{-3}$

http://probaperception.blogspot.com