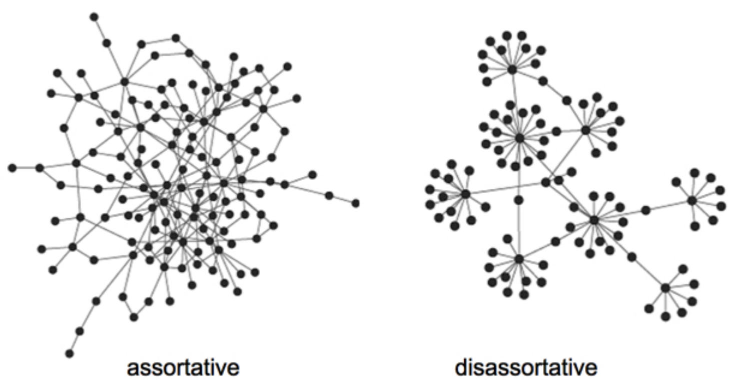
Network Analysis and Modeling
CSCI 5352, Fall 2013
Prof. Aaron Clauset
Problem Set 4, due 10/21

1. (30 pts total) Using the configuration model, investigate the set of random graphs in which all vertices have degree 1 or 3.
   (a) (15 pts) Numerically calculate the mean fractional size of the largest component for a network with $n = 10^4$ vertices, and with $p_1 = 0.6$, $p_3 = 1 - p_1$, and $p_k = 0$ for all other values of k.
   (b) (15 pts) Now make a figure showing the mean fractional size of the largest component for values of p_1 from 0 to 1 in steps of 0.01. Show that this allows you to estimate the value of p_1 for the phase transition at which the giant component disappears.

2. (40 pts) Random graphs with heavy-tailed degree distributions generally exhibit *disassortative* mixing with respect to their degrees, in which high-degree vertices tend to be connected to low-degree vertices (despite the fact that the high-degree vertices tend to link to each other).



   (a) (10 pts) Rewrite Eq. (7.82) in *Networks*, the Pearson correlation coefficient r, to measure degree assortativity as a function of degree, which we denote r_k. Show your work.
   Hint: Recall that a correlation coefficient is related to the covariance over the variance.
   (b) (30 pts) Using the configuration model, numerically investigate the degree assortativity of random graphs with degree distributions given by a power law. To specify the power-law distribution, choose $\alpha = \{1.5, 2.5, 3.5\}$ and let $k \geq 1$. For $n = 10^4$, make a figure (log-linear axes) showing the average behavior of r_k as a function of degree k. Plot trends for all three values of α on the same figure. Briefly summarize how you conducted the experiment and discuss your findings.
   Hint: To create a power-law random graph, you will need to draw degree values from a discrete power-law distribution. This can be done using a numerical version of the *transformation method*, which requires a uniform random deviate r (produced by a pseudorandom number generator) and the cdf of the target distribution. The discrete power-law cdf is $P(x) = 1 - \zeta(\alpha, x)/\zeta(\alpha, x_{\min})$, where $\zeta(\alpha, x) = \sum_{j=x}^{\infty} j^{-\alpha}$ is called the

Hurwitz zeta function. This function is often available in standard numerical libraries, or you may approximate it yourself (it is simply a sum). To transform $r \sim U(0, 1)$ into $x \sim \text{PowerLaw}(\alpha)$, first draw r and then find the smallest x such that $P(x) \geq r$ (more formally we define $x = P^{-1}(r)$, which is the inverse cdf).

3. (30 pts) Using the PS4 network file on the class website, first convert the file into a simple network. Then, compute the mean local clustering coefficient $\langle C \rangle$ as a function of degree k for (i) the empirical network and (ii) a random graph with the same degree sequence. Display your results on a figure, and comment on the similarities or discrepancies.

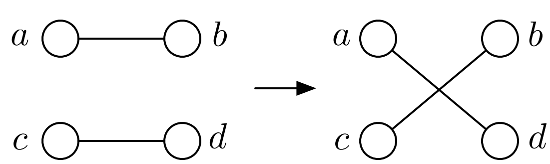
4. (15 pts extra credit) As described in Section 13.2 of *Networks*, the configuration model can be thought of as the ensemble of all possible matchings of edge stubs, where vertex i has k_i stubs. Show that for a given degree sequence, the number Ω of matchings is

$$\Omega = \frac{(2m)!}{2^m m!} \ ,$$

which is independent of the degree sequence.

5. (30 pts extra credit) Using the PS4 network file, first convert the file into a simple network. Then conduct the following numerical experiment on the network's largest component (which contains 379 vertices), which you will need to extract before doing anything else.

   - Compute the eigenvector centrality for each vertex. Call the list of these values v.
   - Now, make a figure showing the Pearson correlation coefficient r for the original eigenvector centrality scores v and the scores v' for a "rewired" network when a fraction $f \in [0, 1]$ (in steps of 0.01) of uniformly random pairs of edges have been rewired by repeatedly applying the following procedure. Choose two edges (a, b) and (c, d) (uniformly at random) and replace them with edges (a, d) and (c, b). Note that this procedure maintains the degree sequence of the graph, but otherwise randomly reorganizes some subset of edges. The trend you are showing is the correlation between v and v' as a function of f, the fraction of edge pairs that have been rewired.
   - Discuss the pattern you observe as you vary f. Suggest a reason for what you see (in terms of the network's structure and the rewiring procedure).



6. (40 pts extra credit) The clustering coefficient for the configuration model, like the coefficient for the classic random graph model, goes to zero in the limit of large n. But for the configuration model, the coefficient's value depends on the degree sequence. Section 13.4 in *Networks* shows that the expected value is

$$C = \frac{1}{n} \frac{\left[\langle k^2 \rangle - \langle k \rangle\right]^2}{\langle k \rangle^3} \ .$$

Using a discrete power-law degree distribution, with choices $\alpha = \left\{\frac{3}{2}, \frac{5}{2}, \frac{7}{2}\right\}$, to generate the degree sequences (see Hint in question 2b above), show via numerical experiments that this expression holds as n increases. Visualize your results on a single figure (log-log axes) that shows the numerical trend and the prediction of the analytic expression above, for each value of α. Briefly summarize how you conducted the experiment, discuss your findings, and offer an explanation for any discrepancies.