Network Analysis and Modeling
CSCI 5352, Fall 2013
Prof. Aaron Clauset
Problem Set 2, due 9/23

1. (15 pts) In a survey of couples in the city of San Francisco in 1992, Catania et al. recorded, among other things, the ethnicity of interviewees and calculated the fraction of couples whose members were from each possible pairing of ethnic groups. The fractions were as follows:

		Women				
		Black	Hispanic	White	Other	Total
Men	Black	0.258	0.016	0.035	0.013	0.322
	Hispanic	0.012	0.157	0.058	0.019	0.246
	White	0.013	0.023	0.306	0.035	0.377
	Other	0.005	0.007	0.024	0.016	0.052
	Total	0.288	0.203	0.423	0.083	

Assuming the couples interviewed to be a representative sample of the edges in the undirected network of relationships for the community studied, and treating the vertices as being of four types—black, hispanic, white, and other—calculate the numbers e_{rr} and a_r that appear in Eq. (7.76) in *Networks* for each type. Hence calculate the modularity Q of the network with respect to ethnicity. What do you conclude about homophily in this community?

2. (35 pts) Suppose a network has a degree distribution that follows the exponential form $p_k = C\mathrm{e}^{-\lambda k}$, where C and λ are constants.

   (a) (4 pts) Find C as a function of λ.

   (b) (5 pts) Calculate the fraction P of vertices that have degrees greater than or equal to k.

   (c) (5 pts) Calculate the fraction W of ends of edges that are attached to vertices of degree greater than or equal to k.

   (d) (6 pts) Hence show that for this degree distribution, the Lorenz curve—the equivalent of Eq. (8.23) in *Networks*—is given by

   $$W = P + \frac{1 - \mathrm{e}^{\lambda}}{\lambda} P \ln P \ . \tag{1}$$

   (e) (5 pts) What is the equivalent of the "80–20" rule for such a network with $\lambda = 1$? That is, what fraction of the "richest" nodes in the network have 80% of the "wealth"?

   (f) (10 pts) Show that the value of W is greater than unity for some values of P in the range $0 \leq P \leq 1$. What is the meaning of these "unphysical" values?

3. (30 pts) Consider the following simple and rather unrealistic model of a network: each of n vertices belongs to one of g groups. The mth group has n_m vertices and each vertex in that group is connected to others in the group with independent probability $p_m = A(n_m - 1)^{-\beta}$, where A and β are constants, but not to any vertices in other groups. Thus, this network takes the form of a set of disjoint groups of communities.

(a) (10 pts) Calculate the expected degree $\langle k \rangle$ of a vertex in group m.

(b) (10 pts) Calculate the expected value $\langle C_m \rangle$ of the local clustering coefficient for vertices in group m.

(c) (10 pts) Hence show that $\langle C_m \rangle \propto \langle k \rangle^{-\beta/(1-\beta)}$. What value would β have to assume for the expected value of the local clustering coefficient to fall off as $\langle k \rangle^{-0.75}$, as has been conjectured by some researchers?

4. (10 pts) A particular network is believed to have a degree distribution that follows a power law. Among a random sample of vertices in the network, the degrees of the first 20 vertices with degree 10 or greater are:

16	17	10	26	13	14	28	45	10	12
12	10	136	16	25	36	12	14	22	10

Estimate the exponent α of the power law and the error on that estimate using Eqs. (8.6) and (8.7) in *Networks*.

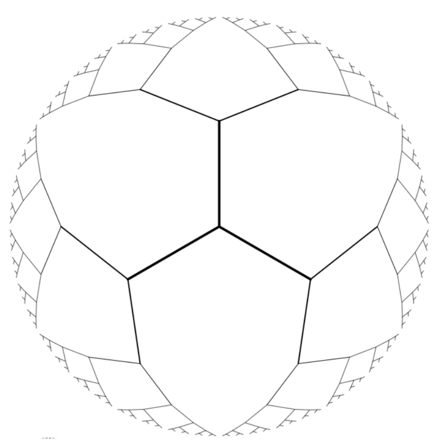
5. (30 pts) Using the PS2 network file on the class website, produce three figures. Remember to label your axes and your data series. For each figure, include a brief interpretation of what it shows relative to what we might expect. For parts (a) and (b), briefly discuss how the patterns for in- and out-degrees contrast.

(a) (10 pts) The complementary cumulative distribution function (ccdf) for the in- and out-degree distributions, on doubly-logarithmic axes, along with maximum-likelihood log-normal distribution fits. Report the parameter estimates. (Hint: when estimating the log-normal parameters, you will need to exclude any vertices with degree $k = 0$.)

(b) (10 pts) The Lorenz curves for the in- and out-degree distributions, on linear axes.

(c) (10 pts) The mean reciprocity $\langle r_k \rangle$ as a function of out-degree k^{out}. (Reciprocity is normally defined as a network average, but the definition can easily be adapted to be a vertex-level measure.)

6. (15 pts extra credit) A Cayley tree is a symmetric regular tree in which each vertex is connected to the same number k of others, until we get out to the leaves, like this, with $k = 3$:



Show that the number of vertices reachable in d steps from the central vertex is $k(k-1)^{d-1}$ for $d \geq 1$. Then give an expression for the diameter of the network in terms of k and the number of vertices n. State whether this network displays the "small-world effect," defined as having a diameter that increases at $\log n$ or slower.