Complex Systems 535/Physics 508: Homework 6

1. Recall the master equations for Price's model of a citation network in the limit of large n:

$$p_q = \frac{c}{c+a}\big[(q-1+a)p_{q-1} - (q+a)p_q\big] \qquad \text{for } q > 0,$$
$$p_0 = 1 - \frac{c}{c+a}p_0.$$

   (i) Write down the special case of these equations for $c = a = 1$.

   (ii) Show that the in-degree distribution generating function $g_0(x) = \sum_{q=0}^{\infty} p_q x^q$ for this case satisfies the differential equation

$$g_0(x) = 1 + \tfrac{1}{2}(x-1)\big[x g_0'(x) + g_0(x)\big].$$

   (iii) Show that the function

$$h(x) = \frac{x^3 g_0(x)}{(1-x)^2}$$

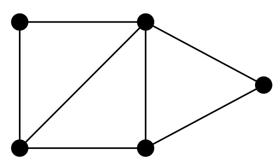
   satisfies

$$\frac{\mathrm{d}h}{\mathrm{d}x} = \frac{2x^2}{(1-x)^3}.$$

   (iv) Hence find a closed-form solution for the generating function $g_0(x)$. Confirm that your solution has the correct limiting values $g_0(0) = p_0$ and $g_0(1) = 1$.

   (v) Thus find a value for the mean in-degree of a vertex in Price's model. Is this what you expected?

2. Consider this small network with five vertices:



   (i) Calculate the cosine similarity for each of the $\binom{5}{2} = 10$ pairs of vertices. (If you need a refresher on cosine similarity, see Section 7.12.1 on page 212 in the book.)

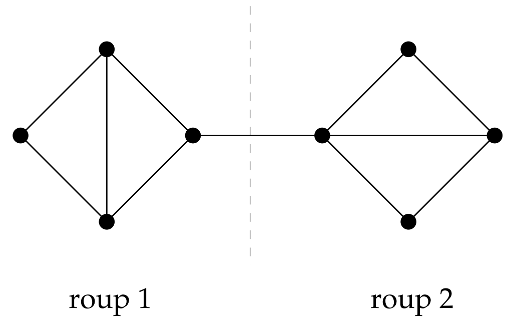
   (ii) Using the values of the ten similarities construct the dendrogram for the single-linkage hierarchical clustering of the network according to cosine similarity.

3. Suppose we draw n random reals x in $[0, \infty)$ from the (properly normalized) exponential probability density $P(x) = \mu \mathrm{e}^{-\mu x}$.

   (i) Write down the likelihood (i.e., the probability density) that we draw a particular set of values x_i (where $i = 1 \ldots n$) for a given value of the exponential parameter μ.

   (ii) Hence find a formula for the best (meaning the maximum-likelihood) estimate of μ given a set of observed values x_i.

4. Consider this small network, divided into two groups as indicated:



(i) Calculate the (three) quantities m_{rs} and the (two) quantities n_r that appear in the profile likelihood for the two-group stochastic block model. Hence calculate the numerical value of the log profile likelihood.

(ii) Verify that no higher profile likelihood can be achieved by moving any single vertex to the other group, and hence that this division into groups is at least a local maximum. (In fact it's the global maximum as well.) Hint: Some of the vertices are symmetry equivalent, which means you need only consider the movement of six different vertices to the other group, which will save you some effort.