Social Network Analysis MATH-517: Homework 4
(Total 10 points - Covers Chapters 10, 11, 12, and 14)

1 [1.5 points] A particular network is believed to have a degree distribution that follows a power law for nodes of degree 10 or greater. Among a random sample of nodes in the network, the degrees of the first 20 nodes 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. (10.9) and (10.10).

2 [1.5 points] Consider the random graph $G(n, p)$ with mean degree c.

a) Show that in the limit of large n the expected number of triangles in the network is $\frac{1}{6}c^3$. This means that the number of triangles is constant, neither growing nor vanishing in the limit of large n.

b) Show that the expected number of connected triples in the network (as defined on page 184) is $\frac{1}{2}nc^2$.

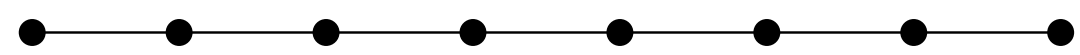
c) Hence calculate the clustering coefficient C, as defined in Eq. (7.28), and confirm that it agrees for large n with the value given in Eq. (11.11).

3 [1.5 points] Consider the random graph $G(n, p)$ with mean degree c.

a) Argue that the probability that a node of degree k belongs to a small component is $(1 - S)^k$, where S is the fraction of the network occupied by the giant component.

b) Thus, using Bayes' theorem (or otherwise) show that the fraction of nodes in small components that have degree k is $\mathrm{e}^{-c}c^k(1 - S)^{k-1}/k!$.

4 [**1.5 points**] Consider a "line graph" consisting of n nodes in a row like this:

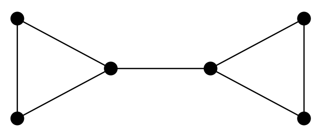


a) Show that if we divide the network into two parts by cutting any single edge, such that one part has r nodes and the other has $n - r$, the modularity, Eq. (7.58), takes the value

$$Q = \frac{3 - 4n + 4rn - 4r^2}{2(n-1)^2}.$$

b) Hence show that when n is even the optimal such division, in terms of modularity, is the division that splits the network exactly down the middle.

5 [**2 points**] Construct the modularity matrix for this small network using Python:



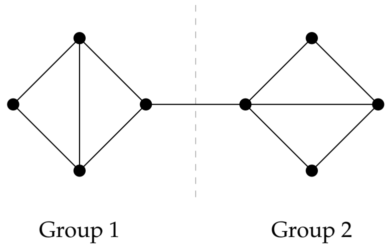
Find the eigenvector of the modularity matrix corresponding to the largest eigenvalue and hence divide the network into two communities.

6 [**2 points**] Write a computer program in the programming language Python that generates a configuration model network with nodes of degree 1 and 3 only and then calculates the size of the largest component.

a) Use your program to calculate the size of largest component for a network of n ⊠ 10 000 nodes with p1 ⊠ 0.6 and p3 ⊠ 0.4 (and pk ⊠ 0 for all other values of k).

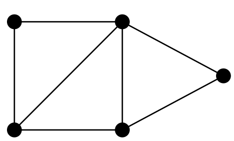
b) Modify your program to calculate the size of the largest component for values of p1 from 0 to 1 in steps of 0.01, then make a graph of the results as a function of p1. Hence estimate the value of p1 at the phase transition where the giant component disappears. Compare your result to the predictions of the analytic calculation in Section 12.6.1.

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



a) For this network calculate the three quantities m_{rs} and the two quantities κ_r that appear in the (log) profile likelihood, Eq. (14.50), for the degree-corrected stochastic block model. Hence calculate the numerical value of the profile likelihood.
b) Verify that no higher profile likelihood can be achieved by moving any single node to the other group, and hence that this division into groups is at least a local maximum of the likelihood. (In fact it's the global maximum as well.) Hint: Some of the nodes are symmetry equivalent, which means you need only consider the movement of six different nodes to the other group, which could save you some effort.

14.7 Consider this small network with five nodes:



a) Calculate the cosine similarity for each of the $\binom{5}{2} = 10$ pairs of nodes. (For a definition of cosine similarity see Section 7.6.1.)
b) Using the values of the ten similarities construct the dendrogram for the single-linkage hierarchical clustering of the network according to cosine similarity.