Complex Systems 535/Physics 508: Homework 4

Because of Fall Break, you have two weeks to do this homework. It is due in class on **Tuesday, October 27**.

1. Consider the random graph $G(n,p)$.
   (i) Show that in the limit of large n the expected number of triangles in the network is $\frac{1}{6}c^3$, where c is the mean degree. In other words, the number of triangles is constant, neither growing nor vanishing in the limit of large n.
   (ii) Show that the expected number of connected triples in the network (as in Eq. (7.41)) is $\frac{1}{2}nc^2$.
   (iii) Hence calculate the clustering coefficient C, as defined in Eq. (7.41), and confirm that it agrees for large n with the value given in Eq. (12.11).

2. Given that the clustering coefficient for $G(n,p)$ goes to zero for large n, can we make a different random graph model that has nonzero clustering coefficient? Indeed we can, as follows. Take n vertices and go through each distinct trio of three vertices, of which there are $\binom{n}{3}$, and with independent probability p connect the members of the trio together using three edges to form a triangle, where $p = c/\binom{n-1}{2}$ with c a constant. You can assume that n is very large.
   (i) Show that the mean degree of a vertex in this model network is $2c$.
   (ii) Show that the degree distribution is $p_k = e^{-c}c^{k/2}/(k/2)!$ if k is even and $p_k = 0$ if k is odd.
   (iii) Show that the clustering coefficient, Eq. (7.41), is $C = 1/(2c+1)$ (which does not go to zero as n becomes large).
   (iv) Show that as a fraction of network size, the expected size S of the giant component, if there is one, satisfies $S = 1 - e^{-cS(2-S)}$.
   (v) What is the value of the clustering coefficient when the giant component fills half of the network?

3. (i) The Fibonacci numbers are 0, 1, 1, 2, 3, 5, 8... They have the definitive property that each is the sum of the previous two. The generating function for the Fibonacci numbers is the power series whose coefficients are the Fibonacci numbers: $f(z) = z + z^2 + 2z^3 + 3z^4 + 5z^5 + \ldots$ Show that $f(z) = z/(1 - z - z^2)$.
   (ii) Consider the binomial distribution of an integer variable k:

   $$B_{nk}(p,q) = \binom{n}{k} p^k q^{n-k},$$

   where $q = 1 - p$. Derive a closed-form expression for the generating function $g(z) = \sum_{k=0}^{n} B_{nk} z^k$. Hence find the mean of the binomial distribution.
   (iii) A sequence of numbers a_k with $k = 1, 2, 3, \ldots$ satisfies the recurrence

   $$a_k = \begin{cases} 1 & \text{for } k = 1, \\ \sum_{j=1}^{k-1} a_j a_{k-j} & \text{for } k > 1. \end{cases}$$

   Show that the generating function $h(z) = \sum_{k=1}^{\infty} a_k z^k = \frac{1}{2}(1 - \sqrt{1-4z})$.

4. Consider the configuration model with exponential degree distribution $p_k = (1 - e^{-\lambda})e^{-\lambda k}$ with $\lambda > 0$, so that the generating functions $g_0(z)$ and $g_1(z)$ are given by Eq. (13.130).

   (i) Show that the probability u in Eq. (13.91) satisfies the cubic equation $u^3 - 2e^{\lambda}u^2 + e^{2\lambda}u - (e^{\lambda} - 1)^2 = 0$.

   (ii) Noting that $u = 1$ is always a trivial solution of Eq. (13.91), show that the nontrivial solution corresponding to the existence of a giant component satisfies the quadratic equation $u^2 - (2e^{\lambda} - 1)u + (e^{\lambda} - 1)^2 = 0$, and hence that the size of the giant component, if there is one, is

   $$S = \tfrac{3}{2} - \sqrt{e^{\lambda} - \tfrac{3}{4}}.$$

   Sketch or plot the form of S as a function of λ.

   (iii) Show that the giant component exists only if $\lambda < \ln 3$.

5. **Extra credit:** Write a computer program in the language of your choice that generates a random graph drawn from the model $G(n, m)$ for given values of n and the average degree $c = 2m/n$, then calculates the size of its largest component. Use your program to find the size of the largest component in a random graph with $n = 1\,000\,000$ and $c = 2\ln 2 = 1.3863\ldots$ and compare your answer to the analytic prediction for the giant component of $G(n, p)$ with the same parameter values. You should find good agreement, even though the models are not identical. (As we discussed in class, the models become essentially the same for large values of m and n because the number of edges in $G(n, p)$ becomes tightly peaked around the expected value $\langle m \rangle = \frac{1}{2}n(n-1)p$.)