Complex Systems 535/Physics 508: Final Examination

Your name:

Please put your name in the space above.

There are seven questions arranged roughly speaking in order of difficulty, with the easiest ones first. The last question is optional, for extra credit—you can get 100% on the test by doing only the first six questions. Each question is worth 10 points. You must show your working to get all the points for a question. You will get partial credit for all portions completed correctly.

You can write your answers on this test paper. If you need more space you can attach extra sheets, but please remember to put your name on those as well. The test is open-book, meaning you can refer to the book, notes you took in class, solutions to homework problems, and so forth. You may not use online resources. You can use a calculator, including graphing calculators.

1. (i) [**5 points**] Describe briefly the type of each of the following networks, e.g., directed, undirected, acyclic, bipartite, tree, planar, etc.

   (a) The worldwide web

   (b) A citation network of papers

   (c) The US Interstate highway network

   (d) A river network

   (e) A network of which teams beat which in college football this season

   (ii) [**5 points**] For each of the following networks describe briefly one empirical technique that could be used to measure the network structure (i.e., to fully determine all the elements of the adjacency matrix):

   (a) The worldwide web

   (b) A citation network of papers

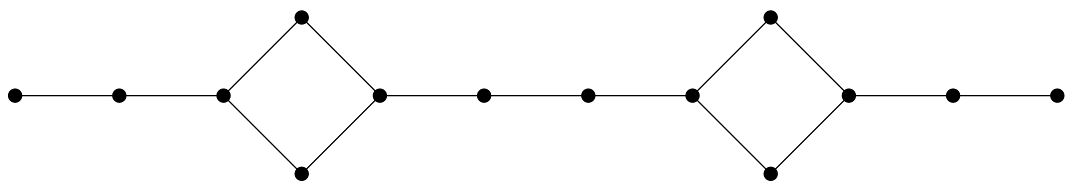
   (c) A food web

   (d) A network of friendships between a group of coworkers

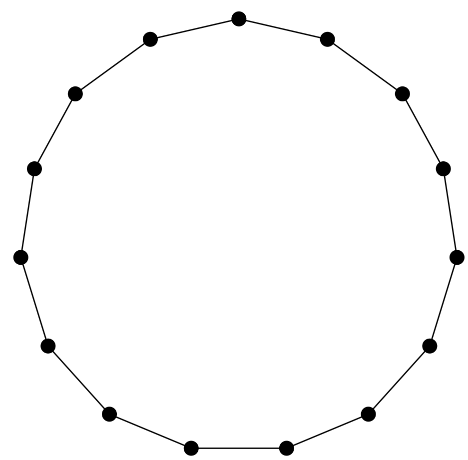
   (e) A power grid

2. (i) [**2 points**] A network is simple and consists of n vertices in a single component. What is the maximum possible number of edges it could have? And what is the minimum possible number of edges it could have? Explain briefly how you arrive at your answers.

(ii) [**2 points**] Circle the 2-core(s) in this graph:



(iii) [**3 points**] Consider a network consisting of n vertices in a ring, where n is odd:



All vertices have the same closeness centrality. What is it, as a function of n?

(iv) [**3 points**] An elementary school class consists of 20 boys and 20 girls. A survey of friendships among the children returns data on 100 friendships, of which 40 are between boys, 40 are between girls, and 20 are between a boy and a girl. What is the modularity?

3. Consider the random graph $G(n, p)$ with mean degree c and consider the bicomponents in that graph (also called 2-components).

   (i) [**2 points**] What fraction of the small components in the network will also be bicomponents?

   (ii) [**3 points**] There can be a giant bicomponent in the network. If there is, it is trivially a subset of the giant component. Moreover, a vertex is in the giant bicomponent if and only if two or more of its neighbors are in the giant component (because the giant component then completes the loop and creates the bicomponent). Recall that the quantity u satisfying $u = \mathrm{e}^{c(u-1)}$ is the probability that a vertex in a random graph is *not* in the giant component. Write an expression for the fraction $1 - S_2$ of vertices in the network that are not in the giant bicomponent in terms of u.

   (iii) [**3 points**] Hence show that $S_2 = S + (1-S)\ln(1-S)$, where S is the size of the giant component.

   (iv) [**2 points**] Thus argue that the random graph contains a giant bicomponent whenever it contains an ordinary giant component.

4. Consider a configuration model with degree distribution p_k.

   (i) [**2 points**] Give expressions for the average number of neighbors and second-nearest neighbors of a randomly chosen vertex. Define any quantities you use in your expressions.

   (ii) [**1 point**] Give an expression for the average number of mth-nearest neighbors of a vertex (assuming n is large)?

   (iii) [**3 points**] What is the average number of neighbors summed over distances 0 to m?

   (iv) [**2 points**] By equating your result with the total number of vertices n in the network, derive an approximate formula for the maximum distance from a vertex to any other.

   (v) [**2 points**] Give two reasons why this result is only approximate.

5. Consider the Barabási–Albert model of a growing network, and define $\pi_k(\tau)\,\mathrm{d}\tau$ to be the fraction of vertices created between times τ and $\tau+\mathrm{d}\tau$ that have degree k, where "time" for the ith vertex added is defined to be $\tau = i/n$, with n being the (current) total number of vertices in the network.

    (i) [**1 points**] What is the value of τ for the most recently added vertex? What is the value for the very first vertex in the network?

    (ii) [**3 points**] Show that
$$\pi_k(\tau) = \binom{k-1}{c-1}\left[\sqrt{\tau}\right]^c \left(1-\sqrt{\tau}\right)^{k-c}.$$
    (You don't need to give a full derivation; you can start from any result given in the book.)

    (iii) [**3 points**] A vertex is observed to have degree k. What is the most likely time at which it was created?

    (iv) [**3 points**] Sketch your result for part (iii) as a function of k for the case $c = 2$ and give a brief explanation in terms of the network's structure and evolution for why it has the shape it does.

6. Consider a uniform random site percolation process on a random 4-regular network (i.e., a configuration model where all vertices have degree 4).

   (i) [**4 points**] Give an expression for the size S of the giant percolation cluster.

   (ii) [**3 points**] Find the critical occupation probability ϕ_c.

   (iii) [**3 points**] Find the value of ϕ at which $S = 1$. This implies that the giant cluster fills the whole network. How can this happen? Surely the most it can fill is the whole of the giant component? Explain.

7. **Extra credit:** Suppose that a search is performed on a peer-to-peer network using the following algorithm. Each vertex on the network maintains a record of the items held by each of its neighbors. The vertex originating a search queries one of its neighbors, chosen uniformly at random, for a desired item and the neighbor responds either that it or one of its neighbors has the item, in which case the search ends, or that they do not. In the latter case, the neighboring vertex then passes the query on to one of *its* neighbors, chosen at random, and the process repeats until the item is found. Effectively, therefore, the search query makes a random walk on the network. (In this simple algorithm there is no prohibition on the walk visiting the same vertex more than once.)

   (i) [**2 points**] Argue that, in the limit of a large number of steps, the probability that the query encounters vertex i on any particular step is $k_i/2m$, where k_i is the degree as usual and m is the total number of edges in the network.

   (ii) [**2 points**] Upon arriving at a vertex of degree k, the search learns (at most) about the items held by all of that vertex's k neighbors, except for the one the query is coming from (which it already knows about), for a total of $k-1$ vertices. Show that on average at each step, in the limit of large number of steps, the search learns about the contents of approximately $\langle k^2\rangle/\langle k\rangle - 1$ vertices and hence that, for a target item that can be found at a fraction c of the vertices in the network, the expected number of copies of the item found on a given step is $c(\langle k^2\rangle/\langle k\rangle - 1)$.

   (iii) [**3 points**] Argue that the probability of not finding the target item on any particular step is approximately $q = \exp[c(1 - \langle k^2\rangle/\langle k\rangle)]$ and that average number of steps it takes to find a copy of the item is $1/(1-q)$.

   (iv) [**3 points**] On a network with a power-law degree distribution with exponent less than 3, so that $\langle k^2\rangle \to \infty$, the results of part (iii) imply that in the limit of large network size the search should end after only one step. Is this really true? If not, why not?