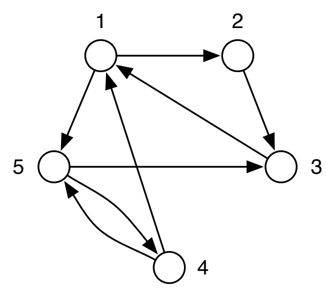
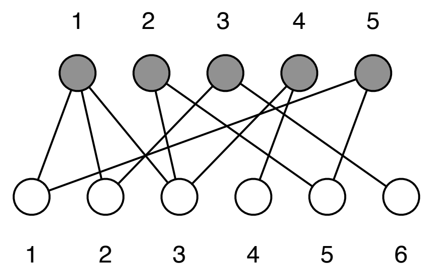
Network Analysis and Modeling
CSCI 5352, Fall 2014
Prof. Aaron Clauset
Problem Set 1, due 9/10

1. (12 pts) Consider the following two networks:



(A)



(B)

   (a) (3 pts) Give the adjacency matrix for network (A).
   (b) (3 pts) Give adjacency list for network (A).
   (c) (6 pts) Give the adjacency matrices for both one-mode projections of network (B).

2. (15 pts) Let **A** be the adjacency matrix of a simple graph (unweighted, undirected edges with no self-loops) and **1** be the column vector whose elements are all 1. In terms of these quantities, multiplicative constants and simple matrix operations like transpose and trace, write expressions for

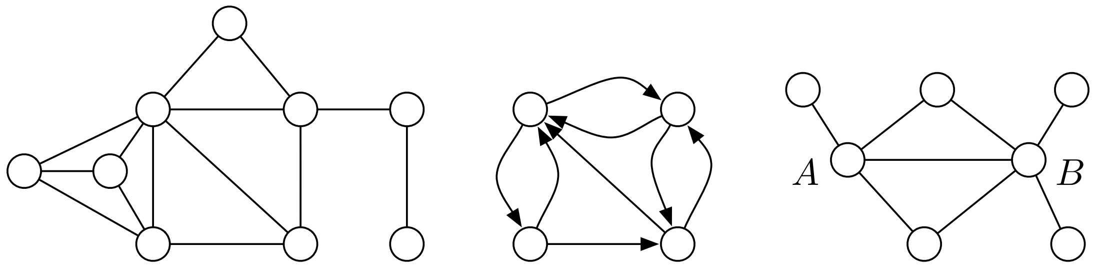
   (a) (3 pts) the vector **k** whose elements are the degrees k_i of the vertices
   (b) (3 pts) the number m of edges in the network
   (c) (5 pts) the matrix **N** whose elements N_{ij} is equal to the number of common neighbors of vertices i and j
   (d) (4 pts) the total number of triangles in the network, where a triangle means three vertices, each connected by edges to both of the others.

3. (10 pts) Consider a bipartite network, with its two types of vertices, and suppose there are n_1 vertices of type 1 and n_2 vertices of type 2. Show that the mean degrees c_1 and c_2 of the two types are given by

$$c_2 = \frac{n_1}{n_2} c_1 \ .$$

4. (13 pts) Consider the following three networks:

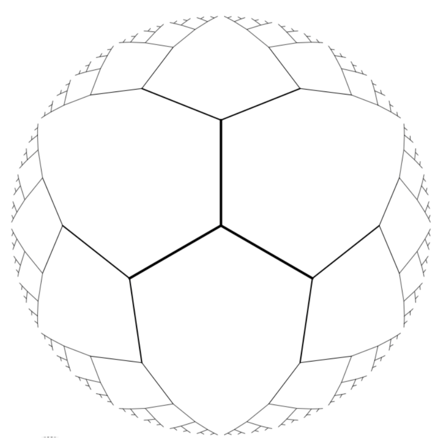
   - (4 pts) Find a 3-core in network (A).
   - (5 pts) What is the reciprocity of network (B)?



(A) (B) (C)

- (4 pts) What is the cosine similarity of vertices A and B in network (C)?

5. (15 pts) 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 the figure on the next page, 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.



6. (35 pts total) Download the Facebook100 ("FB100") data files from the class Dropbox, each of which is a plaintext ASCII file containing an edge list for a 2005 snapshot of a Facebook social network among university students and faculty within some university.[1]

   (a) (20 pts) In many social networks, we observe a surprising phenomenon called the *friendship paradox*. Let k_u denote the degree of some individual u, and let some edge $(u, v) \in E$. The paradox is that the average degree of the neighbor $\langle k_v \rangle$ is *greater* than the average degree $\langle k_u \rangle$ of the vertex. That is, on average, each friend of yours has more friends than you.

[1] The data were kindly provided by A.L. Traud, P.J. Mucha and M.A. Porter, as part of their paper "Social Structure of Facebook Networks," *Physica A* **391**, 4165–4180 (2012), which is freely available at `http://arxiv.org/abs/1102.2166` or `http://bit.ly/1ztbVoS`.

There are two ways to formalize the friendship paradox, as a network-level property or as a vertex-level property. Here, we will consider the vertex-level version, in which the mean neighbor degree (MND) is defined as

$$\langle k_v \rangle = \frac{1}{n} \sum_{u=1}^{n} \left(\frac{1}{k_u} \sum_{(u,v)\in E} k_v \right) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{k_j}{k_i} A_{ij} \ ,$$

that is, for each vertex u, we compute the mean degree of its neighbors, and then we average this quantity across all choices of u.

For each of the FB100 networks, compute and record the mean degree $\langle k_u \rangle$ and the mean neighbor degree $\langle k_v \rangle$. Briefly describe the algorithm you used to do this. Make a figure that plots the ratio $\langle k_v \rangle / \langle k_u \rangle$ (MND over mean degree) versus mean degree $\langle k_u \rangle$. Include a horizontal line representing the line of "no paradox." Comment on the degree to which we observe or do not observe a friendship paradox across these networks. (Figures without axes labels will receive no credit.)

(b) (15 pts) If we assume that edges are distributed randomly, conditioned on each vertex having its observed degree,[2] it can be shown that the network-level expected MND is $\langle k^2 \rangle / \langle k \rangle$, where k is the degree of a vertex. For each of the FB100 networks, compute this quantity. Make a figure that plots the expected network-level MND versus observed vertex-level MND (from part (a)) for the FB100 networks, and includes an appropriate $y = x$ line as a reference frame for interpretation. Comment on the degree to which the two numbers generally agree or disagree (data above or below the $y = x$ line), and explain what these results mean in terms of the vertex-level versus network-level measures.

(c) (20 pts extra credit) Another common property of social networks is that they have very small diameters relative to their total size. This property is sometimes called the "small-world phenomenon" or called having "six degrees of separation".[3]

- For each FB100 network, compute (i) the diameter $\ell_{\max}$ of the largest component of the network and (ii) the mean geodesic distance $\langle \ell \rangle$ between pairs of vertices in the largest component of the network. Make two figures, one showing $\ell_{\max}$ versus network size n and one showing $\langle \ell \rangle$ versus the size of the largest component n. Comment on the degree to which these figures support the six-degrees of separation idea.
- Briefly discuss whether and why you think the diameter of Facebook has increased, stayed the same, or decreased relative to these values, since 2005. (Recall that Facebook now claims to have roughly 10^9 accounts.)

[2] This model is called the *configuration model*, and is a useful family of random graphs, often used as a reference point in network analysis.

[3] This term originated in a play written by John Guare in 1990, which was turned into a 1993 movie starring Will Smith. The concept, however, was originated by the sociologists Stanley Milgram, working in 1967, who was the first to measure the lengths of paths in large social networks.