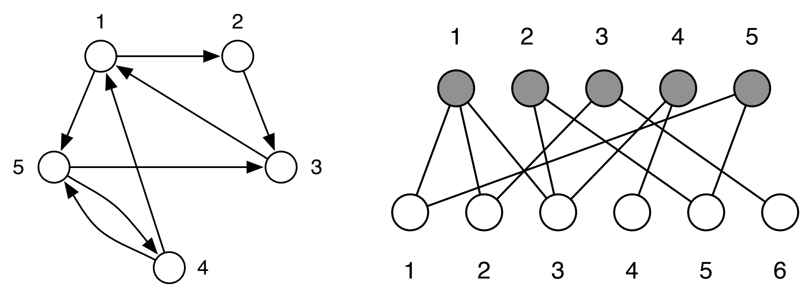
CSCI 5352 Network Analysis and Modeling Fall 2016
Prof. Aaron Clauset
Problem Set 1, due September 6th

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



(A) (B)

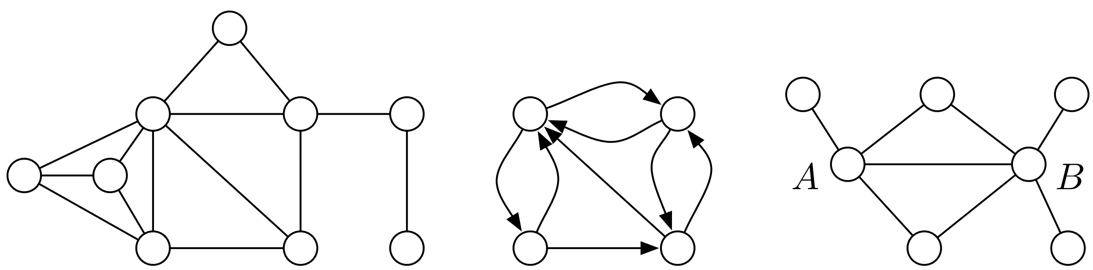
   (a) (3 pts) Give the adjacency matrix for network (A).
   (b) (3 pts) Give adjacency list for network (A).
   (c) (6 pts) Give adjacency matrices for both one-mode projections of network (B).

2. (15 pts) Let $\mathbf{A}$ be the adjacency matrix of a simple graph (unweighted, undirected edges with no self-loops) and $\mathbf{1}$ 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 $\mathbf{k}$ 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 $\mathbf{N}$ 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

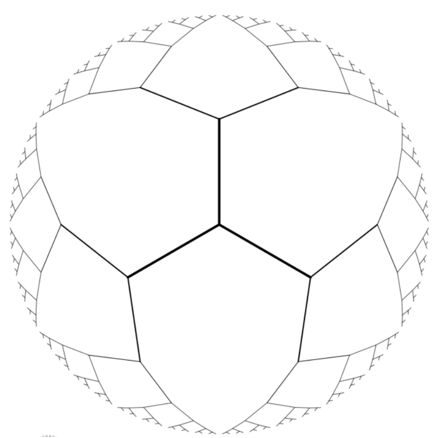
$$c_2 = \frac{n_1}{n_2} c_1 \ .$$

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



(A) (B) (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 below, 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 as $O(\log n)$ or slower.



6. (35 pts total) In this question, we will investigate several properties of online social networks by analyzing the Facebook100 ("FB100") data set, which you may download from the class Dropbox. Each of the 100 plaintext ASCII files in the FB100 folder contains an edge list for a 2005 snapshot of a Facebook social network among university students and faculty within some university.[1]

   (a) (5 pts) In most 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`.

The mean neighbor degree (MND) of a network is defined as

$$\langle k_v \rangle = \frac{1}{2m} \sum_{u=1}^{n} \sum_{v=1}^{n} k_v A_{uv} \quad . \tag{1}$$

Derive an expression for $\langle k_v \rangle$ in terms of the average squared-degree $\langle k^2 \rangle$ and the average degree $\langle k \rangle$. Show your work.

(b) (15 pts) Now, using all 100 of the FB100 networks, make a figure showing a scatterplot of the ratio $\langle k_v \rangle / \langle k_u \rangle$ as a function of the mean degree $\langle k_u \rangle$. Include a horizontal line representing the line of "no paradox," and label the nodes corresponding to Reed, Bucknell, Mississippi, Virginia, and UC Berkeley. (Remember: figures without axes labels will receive no credit.)

Comment on the degree to which we do or do not observe a friendship paradox across these networks as a group. Comment on whether there is any dependency between the size of the paradox (the MND value) and the network's mean degree. A few points of extra credit will be awarded to an explanation of why we should, in fact, expect to see a friendship paradox in these networks, and that identifies the conditions under which we should expect to see *no* paradox.

(c) (15 pts) A related phenomenon in social networks is the *majority illusion.* Let $x \in \{0, 1\}$ be a binary-valued vertex-level property, and let $q = \frac{1}{n} \sum_u x_u$ be the fraction of vertices that exhibit this property. If we set $q < 0.5$, then this property appears only in a minority of nodes. The majority illusion occurs when $q < 0.5$, but the majority of a node's neighbors, on average, exhibit that property, that is, $\langle x_v \rangle > 0.5$.

Explain in words and mathematics how this can be possible.

(d) (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" and is the origin of the popular phrase "six degrees of separation".[2]

- 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 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.