Network Analysis and Modeling
CSCI 5352, Fall 2013
Prof. Aaron Clauset
Problem Set 2, due 9/20

1. (24 pts) Consider the following simple and rather unrealistic model of a network: each of n vertices belongs to one of g groups. The mth group has n_m vertices and each vertex in that group is connected to others in the group with independent probability $p_m = A(n_m - 1)^{-\beta}$, where A and β are constants, but not to any vertices in other groups. Thus, this network takes the form of a set of disjoint groups of communities.

   (a) (8 pts) Calculate the expected degree $\langle k \rangle$ of a vertex in group m.

   (b) (8 pts) Calculate the expected value $\langle C_m \rangle$ of the local clustering coefficient for vertices in group m.

   (c) (8 pts) Hence show that $\langle C_m \rangle \propto \langle k \rangle^{-\beta/(1-\beta)}$. What value would β have to assume for the expected value of the local clustering coefficient to fall off as $\langle k \rangle^{-0.75}$, as has been conjectured by some researchers?

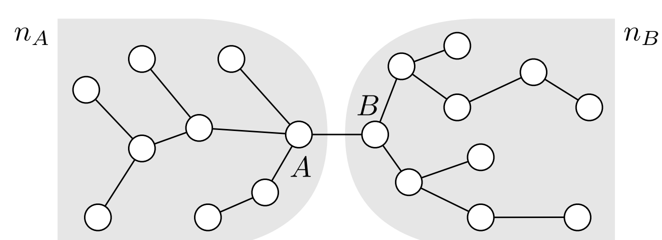
2. (24 pts) Consider the random graph $G(n, p)$ with average degree c.

   (a) (8 pts) Show that in the limit of large n the expected number of triangles in the network is $\frac{1}{6}c^3$. In other words, show that the number of triangles is constant, neither growing nor vanishing in the limit of large n.

   (b) (8 pts) Show that the expected number of connected triples in the network, as in Eq. (7.41) in *Networks*, is $\frac{1}{2}nc^2$.

   (c) (8 pts) Hence, calculate the clustering coefficient C, as defined in Eq. (7.41) in *Networks*, and confirm that it agrees for large n with the value given in Eq. (12.11) in *Networks*.

3. (13 pts) Consider an undirected, unweighted network of n vertices that contains exactly two subnetworks of size n_A and n_B, which are connected by a single edge (A, B), as sketched here:



Show that the closeness centralities C_A and C_B of vertices A and B, as defined by Eq. (7.29) in *Networks*, are related by

$$\frac{1}{C_A} + \frac{n_A}{n} = \frac{1}{C_B} + \frac{n_B}{n} \ .$$

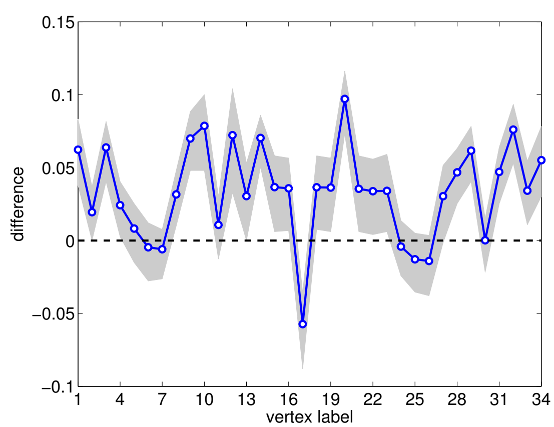
4. (13 pts) Consider an undirected (connected) tree of n vertices. Suppose that a particular vertex in the tree has degree k, so that its removal would divide the tree into k disjoint regions, and suppose that the sizes of those regions are $n_1, \ldots, n_k$. Show that the betweenness centrality b of the vertex is

$$b = n^2 - \sum_{i=1}^{k} n_i^2 \ .$$

5. (26 pts) The Medici family was a powerful political dynasty and banking family in 14th century Florence. The classic explanation, offered by Padgett and Ansell in 1993, of the Medici's rise to power focuses on their establishing themselves as the most central players within the network of prominent Florentine families (shown in Fig. 1; data file in the dropbox). Test their hypothesis by doing the following.

   - Construct a table listing the following centrality scores for each vertex:
     - degree centrality,
     - harmonic centrality (Eq. 7.30 in Networks),
     - eigenvector centrality (Eq. 7.6 in Networks),
     - betweenness centrality[1] (Eq. 7.38 in Networks).

     Within each centrality, sort the (family,value) pairs in decreasing order of importance; three decimal places is sufficient detail. Discuss the degree to which these scores agree with Padgett and Ansell's explanation. And, what do the scores say about the second most important family?
   - Let $\{k_i\}$ be the degree sequence of the network. Produce a figure showing the difference between each vertex's harmonic centrality on the original network and its mean harmonic centrality under the configuration model with $\{k_i\}$. Include the 25 and 75% quantiles around the mean (as in the figure below, for the Karate club). Interpret your results in terms of Padgett and Ansell's story of the Medici family.



[1]To properly compute betweenness, you must handle multiple geodesics with equal length between a pair i and j.

6. (20 pts *extra credit*) Consider the task of designing a network in which distinct vertices hold the status of highest centrality for some set of centrality measures. (Ties are prohibited.) If the network is relatively small compared to the number of centrality measures, this can in fact be a very difficult task.

   - (5 pts *extra credit*) As a warm up, design by hand a small network ($n \leq 7$) that has this property with respect to (i) degree and (ii) betweenness centralities. Include a visualization of the network (label all the vertices, and indicate which ones win at which centrality measure) and produce a table in which the columns show the ranked vertices and scores for these centrality measures.
   - (15 pts *extra credit*) Now, write a program to search the space of all possible networks of a given size n to find a network with this property for (i) degree, (ii) harmonic, (iii) betweenness, and (iv) eigenvector centralities, or report that no such network exists for that size. Run this program for $n = 3, 4, 5, \dots$ and report the smallest value of n for which at least one such network exists. Visualize that network, label all the vertices, and indicate which ones win at which centrality measure; and produce a table in which the columns show the ranked vertices and scores for these centrality measures. Comment briefly about what insights you gained about how to build networks that have distinct winners for different measures of centrality.

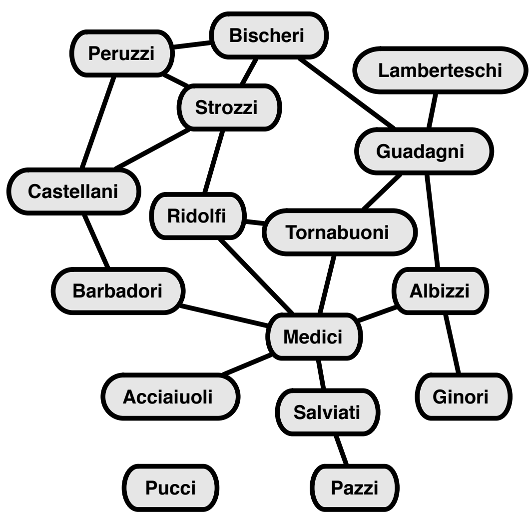


Figure 1: The Medici family alliance network, from Padgett and Ansell (1993).