Homework #1
Networked Life (CIS 112)
Spring 2008
Prof. Kearns

Assigned Feb 12; Due in hardcopy form at the start of class Thursday, Feb 21. This homework should be written up in a separate document in which each answer or solution is clearly labeled.

Problem 1 (10 points)
Recall that a *planar* network is one that can be drawn on a piece of paper without any edges crossing. Draw a diagram of the largest planar network possible that is also a clique (that is, every pair of vertices is connected by an edge).

Problem 2 (20 points)
(a) (10 points) Consider the network of acquaintances within some specific sphere of your life --- for example, the network of acquaintances within your major. Then draw a diagram of your *first neighborhood* within this network --- that is, draw a vertex for yourself, a vertex labeled by each of your acquaintances, and an edge between you and these acquaintances. Then learn which pairs of your neighbors are also acquaintances and draw the corresponding edges. Please be clear about what your network represents (i.e. what sphere you chose and exactly how you defined the edges), and make sure that you chose one in which you have at least four neighbors.
(b) (10 points) Use the network you have drawn in order to compute your own clustering coefficient. Remember that the maximum number of edges possible between K vertices is K(K-1)/2.

Problem 3 (25 points, 5 per part)
For each of the following general categories of networks discussed in class, give a real-world example of such a network other than those given in lecture or the readings. *You should be as clear as possible* about what the vertices are, and about what precisely defines whether there is an edge between two vertices or not. For each of your networks, comment in a sentence or two on its static structure, its dynamics (what takes place on the network), and its formation process (how it grew or formed).
(a) Social Networks
(b) Content Networks
(c) Business or Economic Networks
(d) Physical Networks
(e) Biological Networks

Problem 4 (25 points, 5 points per part)
For this problem you will need to have an account on the Facebook. For those of you who do not already have one, you can obtain one by going to www.thefacebook.com and following the sign-up instructions. Note that in order to complete this problem, there is no need for you to create a publicly viewable profile, nor to have any connections (friends).

(a) Find the individual in the Penn Facebook with the largest number of friends (highest degree) that you can. You can use any method that you like, including those that use information external to the Facebook. Write the name of the individual and the value of their degree, and briefly describe your method of finding them.

(b) Repeat Part (a), but instead find the individual with the *lowest* degree (other than yourself). Again write the name of the individual and their degree, and describe your method of finding them.

(c) By using the Facebook's search facility, and using the "Search All Fields" option, you can find all users who have a chosen word, name, phrase, etc. in their profiles. Pick any phrase such that the resulting search produces at least 30 and at most 100 individuals. Then generate a histogram of the degrees of these individuals. You may use Excel or any other plotting software that you like. Compute and provide the minimum, maximum, and average degrees of this population.

(d) Discuss the extent to which your degree histogram from Part (c) does or does not seem to exhibit the properties discussed in class for such distributions in natural social networks.

(e) A *cycle* in a network is a set of vertices that are connected in a ring; for instance:

A connected to B
B connected to C
C connected to D
D connected to A

where A, B, C and D are all different individuals, is a cycle of length 4. Note that there may or may not be "chords" that cut across this cycle (e.g. B and D might be directly connected). Find the longest cycle you can in the Facebook of length at least 3, and write down the names of the individuals in the order they appear in this cycle. Describe your method for discovering this cycle.

Problem 5 (20 points)

Recall our informal in-class experiments in social navigation, in which the class tried to route the squash ball from an initial person to the final person holding the racquet. In each experiment, person X had "network neighbors" consisting of the people sitting to the left and right of X, and anyone in the class who:

- Experiment A: has the same major as X
- Experiment B: went to high school in the same state as X
- Experiment C: had a cell phone number with the same last digit as X's

The squash ball could only be routed from X to one of his neighbors according to the network structures imposed above.

(a) (10 points) Comment briefly on what you might expect to be the similarities and differences between the three network structures induced by the three rules above, as well as how these structures relate to models we have discussed in lecture and readings. Feel free to conjecture about things like the degree distributions, clustering, etc. as long as you provide some plausible justification.

(b) (10 points) What strategies did the class seem to adopt in trying to solve the navigation problem? What more effective alternate strategies might you suggest?