HW 2: due Thursday, Sept 22

For this assignment, submit your answers as problems as a single pdf file named hwk2.pdf, alongside any code you may have written, networks generated, and required additional text files, using provide with the command:

```
homework% provide comp142 hwk2 p1.csv p2.csv p3.csv hwk2.pdf code-you-wrote
```

Problem 1: For this problem, you will be using three data sets:

- The Karate Club network, a classic example of community structure in networks.
- Dolphins, an undirected social network recording frequent associations between pairs in a community of 62 dolphins living off Doubtful Sound, New Zealand.
- Prison, a directed social network where 67 prison inmates were asked which other inmates they were closest friends with. Each was free to choose as few or as many "friends" as he desired, and "friendships" are not necessarily symmetric.

All three networks are available in the UCINET DL format on the course website.

1. For the karate club network,
   (a) Provide a visualization of the network
   (b) Is the network connected?
   (c) What is the maximum degree?
   (d) Compute the diameter of the network
   (e) Compute the clustering coefficient of the network

(f) Compute the shortest path lengths between each pair of nodes in the network, and provide a histogram of the resulting distribution

(g) Compute the degree for each node in the network, and provide a histogram of the resulting distribution

(h) Compute the betweeness centrality for each node in the network, and provide a histogram of the resulting distribution. Which nodes are most central?

(i) Compute the local clustering coefficient for each node in the network, and provide a histogram of the resulting distribution

(j) Provide your network in the same format that was described for HW1.

2. Repeat part 1 for the Dolphins network.

3. Repeat part 1 for the non-symmetric Prison network. For each measure, state whether or not that measure is meaningful in a non-symmetric network, and compute it if it is.

Note: data available at http://www.cs.tufts.edu/comp/142/private/datasets.html

Problem 2: (This question is due to Jon Kleinberg). In the basic "six degrees of separation" question, one asks if most people in the world are connected by a path of length at most six edges in the social network, where an edge joins any two people who know each other on a first name basis. Now let's consider a variation on this problem. Suppose we consider the entire population of the world, and suppose we create a *directed* edge from each person only to their ten closest friends (but not to other people who they know on a first name basis). In the resulting "closest friend" version of the social network of people in the world, is it possible that for each pair of people in the world, there is a path of length at most six? Explain.