Statistical Analysis of Networks
Statistics 218
Professor: Mark S. Handcock

Homework 1

Due date on the Bruinlearn *Assignments* page

You will need the `networkdata` and `sna` packages. To install the packages from inside `R`:

```
install.packages("network")
install.packages("sna")
install.packages("networkdata",repos="http://faculty.stat.ucla.edu/handcock")
```

1) *Visualization and Sociality:* Here we consider the three network data sets in the `networkdata` package

```
library(networkdata)
data(butland_ppi)
data(addhealth9)
data(tribes)
```

Useful other packages are:

```
library(sna)
library(network)
```

You can check the nature of the networks, via, `help(tribes)`, etc. These networks need to be processed a bit for the analysis to match those reported in class. For the Add Health 9 network, consider only the (known) boys (via `addhealth9$X[,"female"]==0`). Next consider the undirected versions of the networks (For the Add Health 9 network we choose to say there is a tie if either boy nominates the other as a friend. For the tribes take the positive relations. Next consider only the largest component (see, e.g., `component.dist` in `sna`).

a) Create network objects from the networks, using either `network` or `igraph`.

b) Plot the networks so as to visualize their structure.

c) For each network, use the `degree()` function in the `sna` package to find the degree sequence. Note that for undirected networks, this function returns twice the degrees. You may also find the `table` command helpful. *Hint:* The `gmode` option in `degree` is important.

d) For each network, summarize the degree sequence using node-level graphical and numerical summaries (e.g., `barplot`).

2) *Visualizing Connectivity:* Here we consider the Florentine marriage data from the `network` package.

```
library(networkdata)
data(florentine)
```

a) Plot the network, with the names labeled.

b) Which families have the highest and lowest degrees? What are their degrees?

c) Use the `degree()` function in the `sna` package to find the degree distribution. Note that for undirected networks, this function returns twice the degrees. You may also find the `table` command helpful.

d) Which family is excluded from the large component?

3) *Degree distributions:* Degree distributions summarize the densities of ties of the population of nodes. In this question we explore the interactions between proteins of the yeast *S. cerevisiae*. The nodes are types of proteins in the yeast and a directed tie is said to exist if a protein binds to the target protein in a "wet lab" experiment set up to test just this. Not all protein combinations are tested. Here we will consider a series of "mapping" experiments conducted in 2008 that covered approximately 20% of all yeast binary interactions (Yu et. al Science (2008))

a) Go to the home page of the "Yeast Interactome Project":
`http://interactome.dfci.harvard.edu/S_cerevisiae/`
From there download the interactions from CCSB-YI11. These comprise 1809 interactions among 1278 proteins. Construct a `network` object from this edge-list.

b) Construct the out-degree sequence for the network. Construct the in-degree sequence. Are the in- and out-degrees correlated? Use the sum of the in-degree and out-degree as an overall measure of the proteins activity (which we will refer to as its degree).