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

Homework 3

Due date on the Bruinlearn *Assignments* page

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

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

Useful other packages are:

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

1) *Cold war connections*: The dataset `coldwar` contains information on military cooperation and conflict between countries from 1950 through 1985. For this analysis, we will only consider cooperative relations (so all negative relations in the sociomatrix should be set to zero).

```
library(networkdata) data(coldwar)
```

Test the hypothesis that the first military cooperation network (i.e., 1950-1954) follows a simple random graph model (SRG). This is best done with a conditional uniform graph test. Do this using the same degree centralization statistic as in the lectures. Note that this network is undirected, rather than directed as in class.

2) *Mixing within a Classroom:* Here we consider a network of strong friendship ties among 13 boys and 14 girls in a sixth-grade classroom, as collected by Hansell (1984). Each student was asked if they liked each other student "a lot", "some", or "not much". Here we consider a strong friendship tie to exist if a student likes another student "a lot." Also recorded is the sex of each student. The data for the following questions is in the `networkdata` package

```
library(networkdata)
data(hansell)
help(hansell)
```

a) Plot the network. Based on the plot alone, does there appear to be a general preference for same-sex friendship ties?

b) Fit a Erdős-Rényi model, that is, the homogeneous Bernoulli model, to the network using the `ergm` command:

```
library(ergm)
```

```
fit.er <- ergm(hansell ~ edges)
summary(fit.er)
```

Compute, based on the output, the MLE of the log-odds of a tie between two randomly chosen students.

Compute, based on the output, the MLE of the probability of a tie between two randomly chosen students.

How does it compare to the density of ties in the network?

Does the model fit better than the model that says that all networks among these students are equally likely? Quote a statistical test to support your claim.

c) Fit a model for homophily of ties by sex using the `ergm` command:

```
fit.homo <- ergm(hansell~edges + nodematch("sex"))
summary(fit.homo)
```

Based on the output, is their homophily based on sex? Quote a statistical test to support your claim.

Compute, based on the output, the MLE of the log-odds of a tie between two students of the same sex.

Compute, based on the output, the MLE of the probability of a tie between two students of the same sex.

Compute, based on the output, the MLE of the probability of a tie between two students of the opposite sex.

Does the model fit better than the Erdős-Rényi model? Quote a statistical test to support your claim.

d) Fit a model for differential homophily of ties by sex using the `ergm` command:

```
fit.diff.homo <- ergm(hansell~edges + nodematch("sex", diff=TRUE))
summary(fit.diff.homo)
```

Based on the output, does the homophily differ between males and females? Quote a statistical test to support your claim.

e) Fit a model for full mixing between the two sexes using the `ergm` command:

```
fit.mix <- ergm(hansell~edges+nodemix("sex"))
summary(fit.mix)
```

Compute, based on the output, the MLE of the log-odds of a tie from a male to a female student.

Compute, based on the output, the MLE of the probability of a tie from a female to a male student.

How does this compare to the empirical frequency of such ties?

Based on the output, does the full mixing model significantly improve over the other models fit? Quote statistical tests to support your claim.

3) *Centrality and Prestige with the Holland-Leinhardt p_1 model:* Here we consider again Hansen's network of strong friendship ties among 13 boys and 14 girls in a sixth-grade classroom.

Here we consider a network of strong friendship ties among 13 boys and 14 girls in a sixth-grade classroom, as collected by Hansell (1984). Each student was asked if they liked each other student "a lot", "some", or "not much". Here we consider a strong friendship tie to exist if a student likes another student "a lot." Also recorded is the sex of each student. The data are in the `networkdata` package

```
library(networkdata)
data(hansell)
help(hansell)
```

a) Fit the p_1 model using the `ergm` command:

```
fit <- ergm(hansell ~ edges + nodematch("sex") + sender + receiver + mutual)
summary(fit)
```

In this model, does there appear to be a general preference for same-sex friendship ties? Does there appear to be a general preference for mutual friendship ties? Quote statistical tests to support your claims.

Note: There is a bug in `ergm` where the deviance is wrong. To correct it use

```
fit0 <- ergm(hansell ~ edges + nodematch("sex") + sender + receiver +  mutual,
             estimate="MPLE")
fit$mle.lik  <- fit0$mple.lik
```

b) We can interpret the sender coefficients as measures of the "centrality" of the actor, and the receiver coefficients as measures of "prestige".

Plot the prestige verses the centrality scores for each node (see `coef(fit)`). Find on the plot the top two students in terms of prestige and the top two in terms of centrality. Is their top rank apparent in the figure?

How correlated are these measures to (Freeman's) in-degree centrality and out-degree prestige, respectively?

c) Consider the typical magnitude of the sender and receiver coefficients. How do they compare in magnitude to the preferences for same-sex or mutual ties?