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

Homework 4

Due date on the Bruinlearn *Assignments* page

In this homework we consider more sophisticated social network models formed by incorporating structural terms. For a list of the available structural terms and a description of their use type: `help("ergm-terms")`

You may need the `sna` package and the `networkdata` package. You will also need the `ergm.tapered` package, which is open-source, user-friendly software available on GitHub as ergm.tapered: https://github.com/statnet/ergm.tapered

1) *Modeling the French Financial Elite:* Here we consider a network collected by Charles Kadushin and described in the Kadushin (1990).

He collected data from 127 members of the French financial elite. He used various criteria to determine the top 28 and recorded their who-to-whom responses to questions about who was influential, who were members of the elite and who were friends. He also recorded a large amount of information on their individual backgrounds and characteristics.

We will focus on the (undirected) friendship network.

There are many vertex covariates, including:

- prestige: (coded as 0 if respondent has neither a particule nor a social register listing; 1 if a respondent has either a particule or a social register listing; and 2 if respondent has both a particule and social register listing)
- party: An indicator of the party membership. There are 11 parties.
- masons: A member of the masons? 1=no; 2=yes.
- ena: Graduated from ENA? 1=no; 2=yes.
- boards: Number of top boards they are a member of.

Many more are described in the paper.

The data are in the `networkdata` package

```
library(networkdata)
data(ffef)
help(ffef)
```

a) Plot the network with the vertex color being the ENA attendance. What do you see?

b) Fit a model to the network that includes terms for the homophily on ENA attendance, prestige and party affiliation. Include terms for geometrically weighted edgewise shared partners with the scale parameter fixed at 0.5 (i.e., `gwesp(0.5,fixed=T)`). Include a similar term for geometrically weighted dyadwise shared partners with the scale parameter fixed at 0.5 (i.e., `gwdsp(0.5,fixed=T)`).

Hint: Use `ergm.tapered()` as it better deals with strong dependence between terms.

c) Give an interpretation for each of the coefficients in the model in terms of what it means and also what its magnitude indicates about the nature of social relations in the network.

d) Look at the MCMC diagnostics for the model (via, e.g., `mcmc.diagnostics(fit)`). What does it say about the convergence of your model?

e) Extend the model to include other covariates in the network and other terms that you think are interesting in explaining the social structure. Feel free to consult the reference paper for ideas. Overall, what are the important features of the social structure of this network?

2) *Modeling Balance in Friendship Relations:* Here we consider again the network introduced in Homework 3 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 is in the `networkdata` package

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

The statistics `ttriad` and `ctriad` count the number of *transitive triads* and the number of *cyclic triads*, respectively.

a) Is the friendship network balanced in Heider's definition of balance? Give a reason why or why not.

b) We can measure the *statistical* degree of balance in a network by including the `ttriad` and `ctriad` statistics in the model, as these count of the number of *transitive triads* and the count of the number of *cyclic triads*, respectively. If the coefficient of the transitive triad statistic is positive, then the model places higher probability on networks with transitive triads - that is, on balanced networks. We may also see the same or less cyclic triads compared to a neutral random network.

Fit the model with transitive and cyclic triads as well as foundational statistics for the overall density, the mutuality of ties and the homophily on sex using the `ergm.tapered` command:

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

Give a brief interpretation of the coefficients of the first three terms.

Does there appear to be a general preference for mutual friendship ties? Does there appear to be a general preference for same-sex friendship ties?

Quote statistical tests to support your claims.

c) Give a brief interpretation of the coefficients of the `ttriad` and `ctriad` terms.

Describe the pattern of transitive and cyclic ties.

d) Intuitive, what will the coefficient of the transitive triad statistics be if the network is balanced?

Based on this model, does there appear to be a general preference for balanced friendship ties?

3) *Model for Protein-protein interaction data:* Butland et al (2005) "Interaction network containing conserved and essential protein complexes in Escherichia coli" reported a network of protein-protein interactions (bindings) that we obtained from `http://pil.phys.uniroma1.it/~gcalda/cosinsite/extra/data/proteins/`. This data is available in the `networkdata` package

```
library(networkdata)
data(butland_ppi)
help(butland_ppi)
```

Convert the edgelist to a directed network (The `el2sm` function may be helpful).

Fit various tapered ERGM models to the network using `ergm.tapered`. Consider terms documented under `ergm-terms`. Good candidates include `istar`, `ostar`, `gwodegree`, `gwidegree`, `dgwest`, `dgwdsp`, `ctriple`, `ttriple`.

Check the MCMC diagnostics with `mcmc.diagnostics`.

Overall, how does the goodness-of-fit look?