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

Homework 4

Due Friday, November 30, 2018

1) *Modeling a Triad Census 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)
```

a) Fit a triad census model using `ergm.tapered`. Include a term for homophily by sex.

b) Give a brief interpretation of the coefficients of the terms. Based on this model, does there appear to be a general preference for transitive friendship ties? Is there homophily by sex? Give a brief summary of the overall triad census pattern.

2) *Modeling a Triad Census in Friendship Relations:* Here we consider the network of friendship ties girls in grade 9. The data is in the `networkdata` package

```
library(networkdata)
data(gfriends)
help(gfriends)
```

a) First, extract out the grade nine students using `gfriends$X[,1]==9` and create a directed network. Also add a vertex variable equal to the student `gpa`.

Fit a triad census model using a subset of the first seven triad types (012, 102, 021D, 021U, 021C, 111D, 111U). Recall that 003 is the reference category. Also include a GPA, using something like:

```
fit <- ergm.tapered(gf ~ edges + nodecov("gpa")+triadcensus(c(1:7)))
summary(fit)
```

Warning: This will take a few minutes so take a break while it is running!

b) Give a brief interpretation of the coefficients of the terms. Give a brief summary of the overall triad census pattern.

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

d) Try to fit a model with fewer terms than that in a). What is the preferred model?

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?