

Homework 3 — ERGMs

February 12, 2019

1. This question is based on the Network File. There is one object in it — a 40 node network.
 - (a) Provide a summary of important statistics of the network — number of edges, number of two and three stars, number of triangles.
 - (b) Based on the above information what can you say about the network? Comment on how the different statistics relate to each other. What model do you think could have generated this graph?
 - (c) Fit the simplest model: `ergm(g.sim ~ edges)` and check if it adequately fits the other summary statistics. Do so by simulating 1000 networks from the fitted model and plotting the histograms of the three statistics. On each histogram, add a line for the value of the statistic in the original network. Record the AIC and BIC (available in the summary of the model fit or via the AIC and BIC functions).
 - (d) How does the fit look? For example, report the fraction of simulated networks that have values of their statistic that are greater than the observed one. Which of the statistics is fit least well?
 - (e) Fit a new model where you add the statistic that you chose from above to the model in 1.c. Repeat the model fit test.
 - (f) How does the fit look now? Do any other statistics look like they fit better now? Have AIC and BIC improved?
 - (g) Consider the other models that include some or all of the statistics 2-star, 3-star, triangles. Fit them all and report AIC and BIC. What is the best model according to these criteria?

2. ERGMs in action. Simulate 1000 networks of 100 or 1000 nodes each using the following commands (it is now configured to generate 100 node networks, in case 1000 takes too long — changing the size of `g.use2` leads to a change in the size of `g.sim2`):

```
g.use2 <- network(100,density=0.1,directed=FALSE,seed=1)
g.sim2 <- simulate(~edges+kstar(2),
nsim=1000, coef=c(-1,1),
basis=g.use2,
control=control.simulate(
MCMC.burnin=1000,
MCMC.interval=100),seed=2)
```

Count the number of edges and number of 2-stars in each of those graphs and plot these values (try a histogram and maybe a traceplot — the sampling is being done via MCMC after all). What can you say about the graphs you simulated? Why do you think this is happening?