Homework — multiple due dates!

Alexander Volfovsky

March 28, 2019

1. Due Tuesday, April 2: one page update about where you are with the project. It should include a brief literature review and exploratory data analysis/preliminary results.

2. Due Thursday, April 4: Derive the covariance information for different dyads in the following model:

$$\begin{aligned} y_{ij} =& \mu + a_i + b_j + \epsilon_{ij} \\ (a_i, b_i) \overset{iid}{\sim}& N(0, \Sigma_{ab}) \\ (\epsilon_{ij}, \epsilon_{ji}) \overset{iid}{\sim}& N(0, \Sigma_{\epsilon}) \end{aligned}$$

   Specifically consider $cov(y_{ij}, y_{ij})$, $cov(y_{ij}, y_{st})$, $cov(y_{ij}, y_{ik})$, $cov(y_{ij}, y_{jk})$ and $cov(y_{ij}, y_{ki})$.

3. Due Tuesday, April 9: Exploring the latent space. Download the data from the course website and read it into R; you should have a binary adjacency matrix labeled "Y", covariates: a vector of GPAs, a vector of ages and a matrix labeled "nemails" that captures the number of emails sent from person i to person j. We are interested in understanding the probabilities of friendships among these people, possibly as a function of the provided covariates. (Whenever you fit a model, include its output)

   - Perform some exploratory data analysis on the network: compute/plot average in/out degrees; compare those (and/or other statistics) to the covariates. Suggest a possible model for y_{ij} given the covariates and maybe some latent effects. Consider constructing other dyadic covariates using the unit-level effects (differences?).
   - Fit a baseline model:

```
ame0 <- ame(Y,rvar=FALSE,cvar = FALSE, dcor = FALSE,model="bin")
```

and comment on the goodness of fit statistics.

- Fit a weird baseline model with multiplicative effects:

  ```
  ame0_r5 <- ame(Y,rvar=FALSE,cvar = FALSE, dcor = FALSE,model="bin",R=5)
  ```

  and comment on the goodness of fit statistics. What might be happening here?

- Explore some models including the given covariates and constructed covariates. Provide a justification for the selected models and interpret the parameters. Specifically, for each model you choose to describe (lets call that model object "mod") describe the posterior estimates of the βs:

  ```
  apply(mod$BETA,2,function(x)quantile(x,c(.025,.5,.975)))
  ```

- Pick your favorite model. If you did not use any random effects (a, b, u, v) justify your choice. If you used any of them, plot their posterior means (for example, you might want to plot the posterior means of a versus b, the posterior means of the first two dimensions of u against each other, etc). When making these plots also consider coloring the points according to some of the observed covariates—does it look like there is a relationship between the latent space and the observed covariates? If there is, what is it? If there isn't, what does that mean about the latent space?

- Lastly: fit some other model—use "ergm", "netlm" or any other package (though do make sure you know what the package is doing). Comment on the differences and similarities in the results of the model.