

Social Network Analysis – Winter 2022

Lab 3: Autologistic Actor Attribute Model (ALAAM)

Deadline: February 25th, 11:59 pm (CT)

Deliverables: a single PDF file. Please include all the plots and needed tables in this file.

In this lab, you will be building, estimating, and interpreting the results of Autologistic Actor Attribute Model (ALAAM) using the R to examine the effects of network ties on certain behaviors. The R script is based on Johan Koskinen's workshop material and recent paper on Bayesian Analysis of Social Influence¹. In this code, they extended on current approaches for fitting ALAAMs by presenting a comprehensive Bayesian inference scheme that supports testing of dependencies across subsets of data and the presence of missing data. Please do NOT disseminate the code without permission.

Part I: Data Description

You will be using data from the *Teenage Friends and Lifestyle Study*². The provided R code will download the data from https://www.stats.ox.ac.uk/~snijders/siena/s50_data.zip. The dataset for this lab assignment includes a network containing friendship relationships between 50 teenage girls recorded at a point in time.

1. s50-network1.dat

The data also includes information about the attributes of people as follows. We use three of the binary variables for this lab assignment as follows.

2. Binary variables are the binary attributes of the students:
 - 2.1. The *exercising behavior (sport)* of the 50 female students in s50-sport.dat file. 1 (regular), 0 (not regular).
 - 2.2. The *smoking behavior* of the 50 female students in s50-smoke.dat file. The smoking variable has two levels: 1 (smokes regularly), 0 (otherwise).
 - 2.3. The *alcohol consumption* of the 50 female students in s50-alcohol.dat file. 1(regular user), 0(nonregular user)

We use smoking behavior as the dependent variable in our models. Therefore, we are interested in predicting smoking behavior based on the friendship network and other attributes.

Part II: Hypotheses

In this lab assignment you will test the following hypotheses.

¹ Koskinen, J., & Daraganova, G. (2020). Bayesian Analysis of Social Influence. *arXiv preprint arXiv:2006.16464*.

² For more details, check this website https://www.stats.ox.ac.uk/~snijders/siena/s50_data.htm

When contagion effect is not present in the model:

Hypothesis 1: Those who use alcohol regularly, are more likely to smoke regularly.

Hypothesis 2: Those who are reported as friends more, are more likely to smoke regularly.

Hypothesis 3: Those who report more friends, are more likely to smoke regularly.

Hypothesis 4: Those who have more reciprocal friends, are less likely to smoke regularly.

Hypothesis 5: Those who are reported as friends by many people, are less likely to smoke regularly.

When simple contagion effect is present in the model:

Hypothesis 6: There is a positive contagion effect for smoking behavior.

Hypothesis 7: Those who exercise regularly, are more likely to smoke regularly.

Hypothesis 8: Those who use alcohol regularly, are more likely to smoke regularly.

Hypothesis 9: Those who report more friends, are more likely to smoke regularly.

Part III: Initializing the Model

1. First you define and calculate the covariates for all the nodes of the network and put them in a matrix called “covs”. We intend to use all or a subset of the following covariates in our ALAAM models.

Attribute variables:

Sport: exercise activity attribute of nodes

Alcohol: alcohol attribute of nodes

Model terms:

in.degree : number of ties received

out.degree : number of ties sent

rec.ties : number of mutual ties

in.two.star : in-stars reflecting dispersion in popularity

out.two.star : out-stars reflecting dispersion in activity

mix.two.star : correlation between indegree and outdegree (two-path)

in.three.star : further measure of in-degree heterogeneity

out.three.star : further measure of out-degree heterogeneity

triangles : embedded in transitive triads

Part IV: Building the Models

A. Model 1 (non-contagion model):

A.1. (2 points) Build the first ALAAM model by using “BayesALAAM” function taking smoking behavior as the dependent variable. Include all the covariates you built in previous section. Also set “contagion = 'none'” to have a non-contagion model. Set the number of iterations to 1,000. Call this model res.0. Include the final table that you get after running the model in your report.

A.2. (5 points) You will notice in the output that the simple contagion effect is reported as zero because it hasn't been estimated. From the table you got in (A.1) you see the effective sample sizes (ESS). What those numbers mean to you? What are good values for ESS in general?

A.3. (7 points) Plot the MCMC output in trace plots and include them in your report. What those plots tell you? (i.e. How those plots supposed to be? Are they as supposed to?)

A.4. (14 points) Now increase the number of iterations to 10,000 and run the model again and include the table in your report. Again, plot the MCMC output in trace plots and include them in your report. Compare the model that you ran in part (A.1) with this model in terms of ESS and MCMC plots. Is the model with more iterations improved?

A.5. (26 points) Using “write.res.table” function, summarize the results of the model with 10,000 iterations and include the table in your report. Using the table decide on the Hypotheses 1 to 5. Note that for each hypothesis you should say if the hypothesis is supported or not and why is that. (Hint: look at the sign of the mean. Also note that if 0 is included in the interval [the last two columns], the parameter test is not significant.)

B. Model 2 (simple social contagion model):

B.1. (2 points) Build the simple social contagion ALAAM model by using “BayesALAAM” function taking smoking behavior as the dependent variable. Include Sport, alcohol and outdegree from the list of covariates you built in previously. Also set “contagion = 'simple'” or just simply not include that (simple contagion is the default of the function). By doing so, you have a simple contagion model. Set the number of iterations to 1,000. Call this model res.1. Include the final table that you get after running the model in your report.

B.2. (7 points) Plot the MCMC output in trace plots and include them in your report. What those plots tell you? (i.e. How those plots supposed to be? Are they as supposed to?)

B.3. (2 points) Improve the model with taking the theta estimates from the model as inputs to another model. It is possible by setting Propsigma to thetas in BayesALAAM function. Using this trick build a model as in B.1 but set the Propsigma. Also increase the iterations to 5,000. Call the model res.2 and run the model again and include the table in your report.

B.4. (7 points) Plot the MCMC output in trace plots for the improved model and include them in your report. How those changed as compared to the ones in B.2?

B.5. (21 points) Using “write.res.table” function, summarize the results of the model res.2 and include the table in your report. Using the table decide on the Hypotheses 6 to 9. Note that for each hypothesis you should say if the hypothesis is supported or not and why is that. (Hint: look at the sign of the mean. Also note that if 0 is included in the interval [the last two columns], the parameter test is not significant.)

Part V: Goodness-of-fit test

1. **(7 points)** Based on the posterior draws in res.0 model with 10,000 iterations (Thetas of the model), draw outcomes for goodness-of-fit for model 1 and put the last table in your report. Interpret the gof results. (Hint: higher p-values mean better fit).