Lab 3 - ALAAM (Autologistic Actor Attribute Model)

CompSci 396-0: Social Networking Analysis ***Win 2022***

Student Name: Jiaqi Guo NetID: JGR9647

● Responses to Question

■ Part II: Hypotheses

Contagion effect not presented:

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.

Contagion effect presented:

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.

■ Model A: Non–contagion Model

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.

```
summaries of the posterior draws:
                    mean         sd        ESS    SACF 10    SACF 30
intercept      -3.7337001  2.6684208 19.4987848  0.6275812  0.2767491
contagion       0.0000000  0.0000000  0.0000000
Sport          -2.3742551  1.1087631 27.0939297  0.5831772  0.1756092
Alcohol         2.2760708  0.5615391 24.0847168  0.5321850  0.1178511
indegree       -2.1986763  2.1996860 20.5769698  0.6851720  0.4343417
outdegree       1.8236303  2.0558283 12.8168327  0.7769267  0.6051408
reciprochation -0.3120603  1.0494113 17.2992145  0.7064724  0.2731831
instar         -2.4158141  1.8972380 12.7471490  0.7831358  0.6372561
outstar         5.5561658  2.1779521 20.4833300  0.7187569  0.3728744
twopath        -0.6694262  0.4786654 31.8760507  0.4967769  0.1088435
in3star         1.6205537  0.9545902  7.2415602  0.8844524  0.7961092
out3star       -3.0361300  1.3339278 18.6064193  0.7264343  0.3595339
transitive     -0.9458183  0.6202302 19.4359114  0.6624939  0.3572159
```

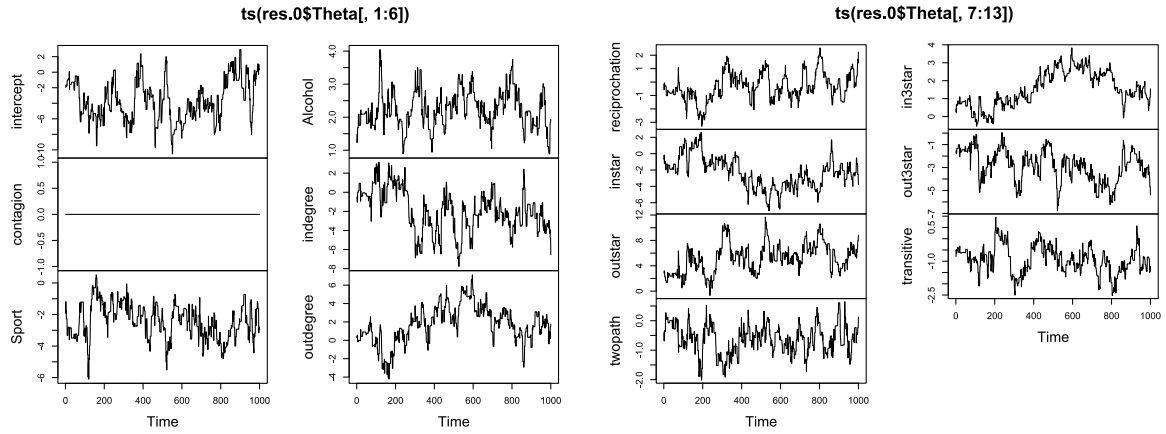
2. (5 points) You will notice in the output that the simple contagion effect is reported as zero because it hasn't been estimated. From the table you got in (A.1) you see the effective sample sizes (ESS). What those numbers mean to you? What are good values for ESS in general?

The effective sample size (ESS) is calculated by:

$$ESS = \frac{n}{1 + 2\sum_{k=1}^{\infty} \rho(k)}$$

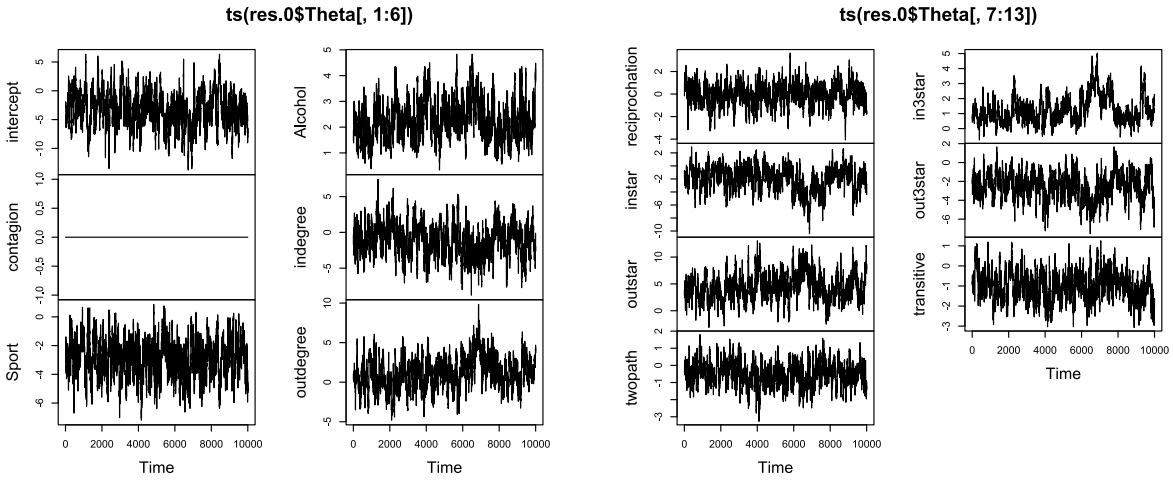
The higher the ESS value, the more information we can get from the simulation. Therefore, a ESS value >100 would be good, a ESS value >200 would be better. But chasing ESS value >100 may be a waste of computation resources it is not worth it.

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?)



Those plots are supposed to be more stabilized around a certain value to arrive to their desired distribution. As shown above, we can observe random fluctuations in almost all the plots (does not converge), so those estimates are not so satisfactory.

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?



Compared to the plots in A.1, the plots here are fluctuate*(follow normal distribution) around a certain value, that is, our model converge to a certain value. So, the model does improve with more iteration applied.

5. **(5 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.)**

	parameter	mean	sd	.025	0.975
1	intercept	-3.163	3.041	-9.236	2.938
2	contagion	0.000	0.000	0.000	0.000
3	Sport	-2.886	1.348	-5.596	-0.349
4	Alcohol	2.267	0.742	0.935	3.903
5	indegree	-1.157	2.318	-5.601	3.419
6	outdegree	1.148	2.006	-2.812	5.104
7	reciprochation	-0.062	0.966	-1.959	1.880
8	instar	-2.013	1.873	-6.064	1.178
9	outstar	4.410	2.569	-0.294	9.603
10	twopath	-0.496	0.694	-1.813	0.885
11	in3star	1.255	0.902	-0.069	3.487
12	out3star	-2.509	1.431	-5.520	-0.022
13	transitive	-0.986	0.703	-2.332	0.419

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

For parameter 4, alcohol, the 95% confidence interval [0.935, 3.903] does not across 0, and it is positive estimate, which suggests that those who use alcohol regularly are 9.65 times more likely to smoke regularly. Therefore, this hypothesis is supported.

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

For parameter 5, indegree, the 95% confidence interval [-5.601, 3.419] does across 0(not statistically significant), and the estimate is positive, which means this hypothesis is not supported.

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

For parameter 6, outdegree, the 95% confidence interval [-2.812, 5.104] does across 0(not statistically significant), and the estimate is positive, which means this hypothesis is not supported.

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

For parameter 7, reciprocation, the 95% confidence interval [-1.959, 1.880] does across 0(not statistically significant), and the estimate is negative, which means this hypothesis is not supported.

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

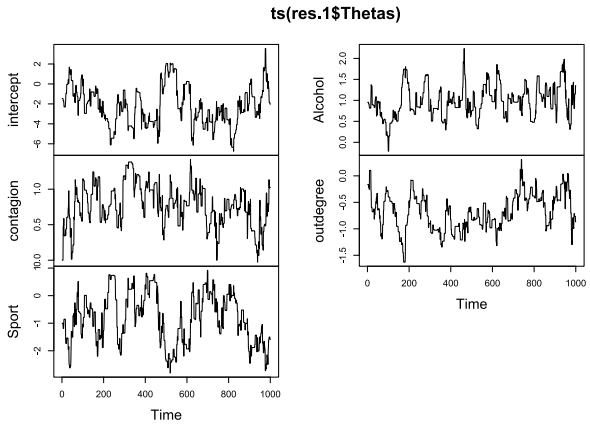
For parameter 8, instar, the 95% confidence interval [-6.064, 1.178] does across 0(not statistically significant), and the estimate is negative, which means that this hypothesis is not supported

■ Model B: Simple Social Contagion Model

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.**

```
summaries of the posterior draws:
                mean          sd        ESS     SACF 10     SACF 30
intercept -1.49258358  1.72271722 25.26621824  0.54751071  0.05425233
contagion  0.66862821  0.32736384 20.06111124  0.57552300  0.15134449
Sport     -1.03367134  0.99734106 14.67990056  0.68916157  0.27192730
Alcohol    0.95173326  0.42042940 16.38970939  0.68476033  0.35738690
outdegree -0.68566666  0.28944690 18.86568676  0.67255196  0.31213256
```

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?)**

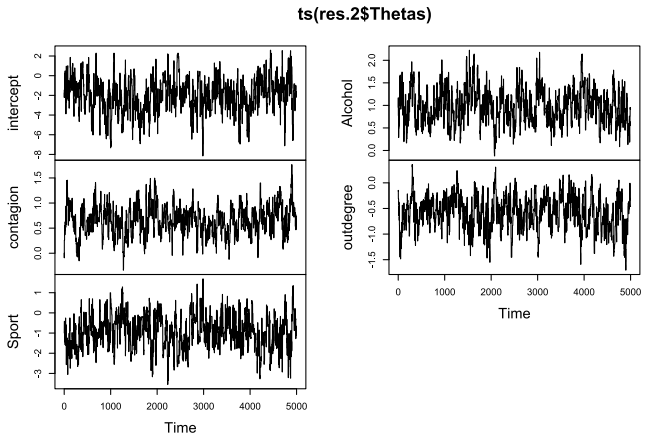


Those plots are supposed to be more stabilized around a certain value to arrive to their desired distribution. As shown above, which suggests that we can observe random fluctuations (not symmetry) in almost all the plots (dose not converge), so those estimates are not so satisfactory.

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**

```
                mean          sd        ESS     SACF 10     SACF 30
intercept  -2.0255788   1.7253834 145.2448251   0.5377802   0.1974450
contagion   0.6618515   0.2862721 118.7446959   0.6159602   0.2431267
Sport      -0.8862288   0.8144669 143.9736101   0.5527777   0.2126368
Alcohol     0.9664831   0.3684623 129.3976362   0.5461028   0.1939791
outdegree  -0.5781710   0.3218781 125.9530998   0.5746021   0.2195389
```

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?**



Compared to the plots in B.2, the plots here are fluctuate around a certain value*(follow normal distribution), which suggests that our model converge to a certain value. So, the model does improve by setting *Propsigma* and increasing the iteration.

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.)**

	parameter	mean	sd	.025	0.975
1	intercept	-1.933	1.711	-5.732	1.375
2	contagion	0.657	0.293	0.086	1.273
3	Sport	-0.929	0.814	-2.545	0.555
4	Alcohol	0.971	0.373	0.291	1.732
5	outdegree	-0.582	0.322	-1.236	-0.001

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

For parameter 2, contagion, the 95% confidence interval [0.086, 1.273] does not across 0, and it is positive estimate, which suggests that a person is 1.92 times more likely to smoke if they are around people who also smoke. Therefore, this hypothesis is supported.

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

For parameter 3, sport, the 95% confidence interval [-2.545, 0.555] does across 0(not statistically significant), and the estimate is negative, which means this hypothesis is not supported.

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

For parameter 4, alcohol, the 95% confidence interval [0.291, 1.732] does not across 0, and it is positive estimate, which suggests that a person is 2.64 times more likely to smoke regularly if they are also alcohol regularly. Therefore, this hypothesis is supported.

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

For parameter 3, outdegree, the 95% confidence interval [-1.236, -0.001] does not across 0, and it is negative estimate, which suggests that a person is 0.55 times less likely to smoke if they are reported more friends. Therefore, this hypothesis is not supported.

■ Part V: Goodness-of-fit test

1. (10 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).

	obs	mean	p-val
intercept	17.000	17.730	0.219
simple cont.	25.000	24.598	0.187
recip cont.	7.000	8.520	0.176
indirect cont.	59.000	42.862	0.102
closedind cont.	22.000	21.104	0.189
transitive cont.	13.000	15.776	0.196
indegree	43.000	42.070	0.196
outdegree	40.000	42.518	0.214
twopaths	26.000	28.514	0.197
out2star	57.000	55.142	0.255
in2star	44.000	40.176	0.171
outtria	88.000	86.614	0.230
intria	71.000	55.152	0.238
transtri	22.000	19.454	0.184
indirecct ties	34.000	34.270	0.265

If p-value is greater than 0.05, it indicates that our simulation is a good fit. As shown in the figure above all the p-value is above 0.05, which suggests that our simulated model can fit the observed data very well.