Part I: Data Description

Binary variables are the binary attributes of the students:

1. The exercising behavior (sport) of the 50 female students in s50-sport.dat file.
   1 (regular), 0 (not regular).
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).
3. The alcohol consumption of the 50 female students in s50-alcohol.dat file.
   1(regular user), 0(nonregular user)

Part II: Hypotheses

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

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):

1. **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.**

Answer:

```
summaries of the posterior draws:
                     mean          sd        ESS     SACF 10     SACF 30
intercept      -2.53774041  2.95622284 14.55906169  0.69980746  0.32401768
contagion       0.00000000  0.00000000  0.00000000
Sport          -2.87334958  1.32805975 19.61077266  0.62964326  0.01770163
Alcohol         2.11842456  0.70434625  9.61134594  0.74659477  0.52074925
indegree       -0.77637375  2.29855932 13.52808903  0.72828183  0.43654399
outdegree       1.12728835  2.07645955 14.07278683  0.75041750  0.46948998
reciprochation -0.39140900  0.96747163 20.35355925  0.66131436  0.31881160
instar         -1.98205231  1.82890603 14.52576549  0.75832157  0.50340154
outstar         4.08095010  2.22851740 14.59255937  0.68631798  0.44087586
twopath        -0.49190006  0.55146896 29.50349100  0.48832582  0.14166272
in3star         1.26794098  0.91754154  7.33530316  0.86598203  0.68677866
out3star       -2.28659494  1.21863118 16.87018784  0.69279539  0.44835166
transitive     -0.99883576  0.58020759 27.16728430  0.46739703 -0.01872642
```

2. **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?**

Answer:

ESS is the Effective Sample Size.

ALAAM draws networks samples from a Markov Chain Monte Carlo(MCMC) sample.

In statistics, ESS is a heuristic to measure how many of these MCMC samples are independent.

- If all samples are correlated, the ESS = 1. (One sample is worth it).
- If all samples are not correlated, ESS = 1000(iteration number). (All the samples are worth it).

Calculation:

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

where n is the number of samples and $\rho(k)$ is the correlation at iteration k

- If all samples are independent($p = 0$) , then ESS = n
- If all samples are dependent($p = 1$), then ESS ~ 1

In general, the larger is better. For example, if we take 1000 as number of samples. Greater than 100 will be good, greater than 200 will be better. However, if we chase ESS more than 10000, it may be a waste of computational resources.

3. **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?)**

Answer:
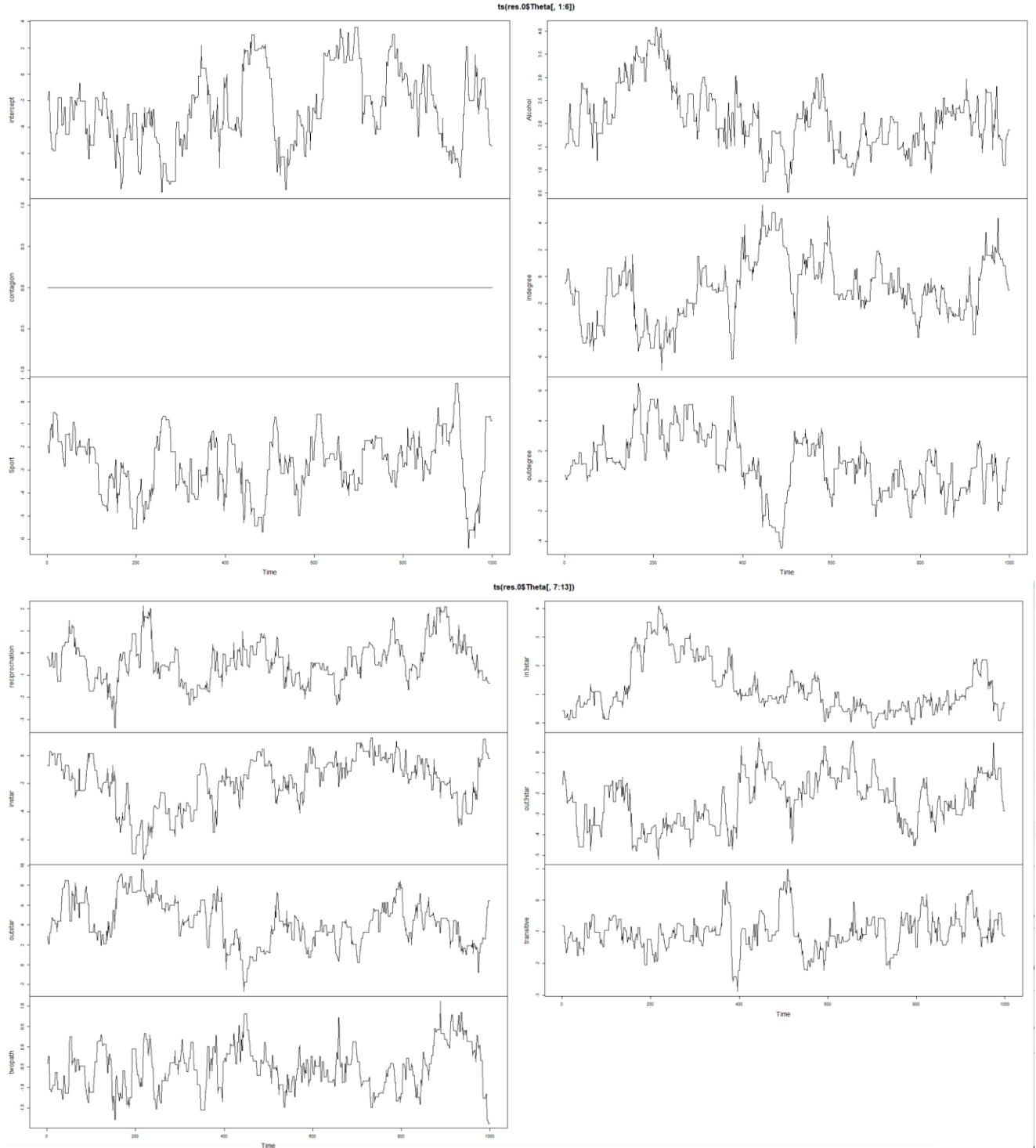
Basically, we could observe from the MCMC samples:

1. **Intercept** randomly fluctuates around the mean value(-2.53). The graph is generally fit a normal distribution centered around -2.53.
2. **Sport** randomly fluctuates around the mean value(-2.87). The graph is generally fit a normal distribution centered around -2.87.

3. **Alcohol mean value is 2.12,** its graph here is abnormal, the first half is generally greater than the second half, it is not a good distribution, and it is supposed to be normal distribution.

4. **indegree** randomly fluctuates around the mean value(-0.77). The graph is generally fit a normal distribution centered around -0.77.

5. **outdegree mean value is 1.13,** its graph here is abnormal, the first half is generally greater than the second half, it is not a good distribution, and it is supposed to be normal distribution.

6. **reciprochation** randomly fluctuates around the mean value(-0.39). The graph is generally fit a normal distribution centered around -0.39.

7. **instar** randomly fluctuates around the mean value(-1.98). The graph is generally fit a normal distribution centered around -1.98.

8. **outstar mean value is 4.08,** its graph here is abnormal, the first half is generally greater than the second half, it is not a good distribution, and it is supposed to be normal distribution.

9. **twopath** randomly fluctuates around the mean value(-0.49). The graph is generally fit a normal distribution centered around -0.49.

10. **in3star** randomly fluctuates around the mean value(1.26). The graph is generally fit a normal distribution centered around 1.26.

11. **out3star** randomly fluctuates around the mean value(-2.28). The graph is generally fit a normal distribution centered around -2.28.

12. **transitive** randomly fluctuates around the mean value(1.00). The graph is generally fit a normal distribution centered around 1.00.

4. **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?**
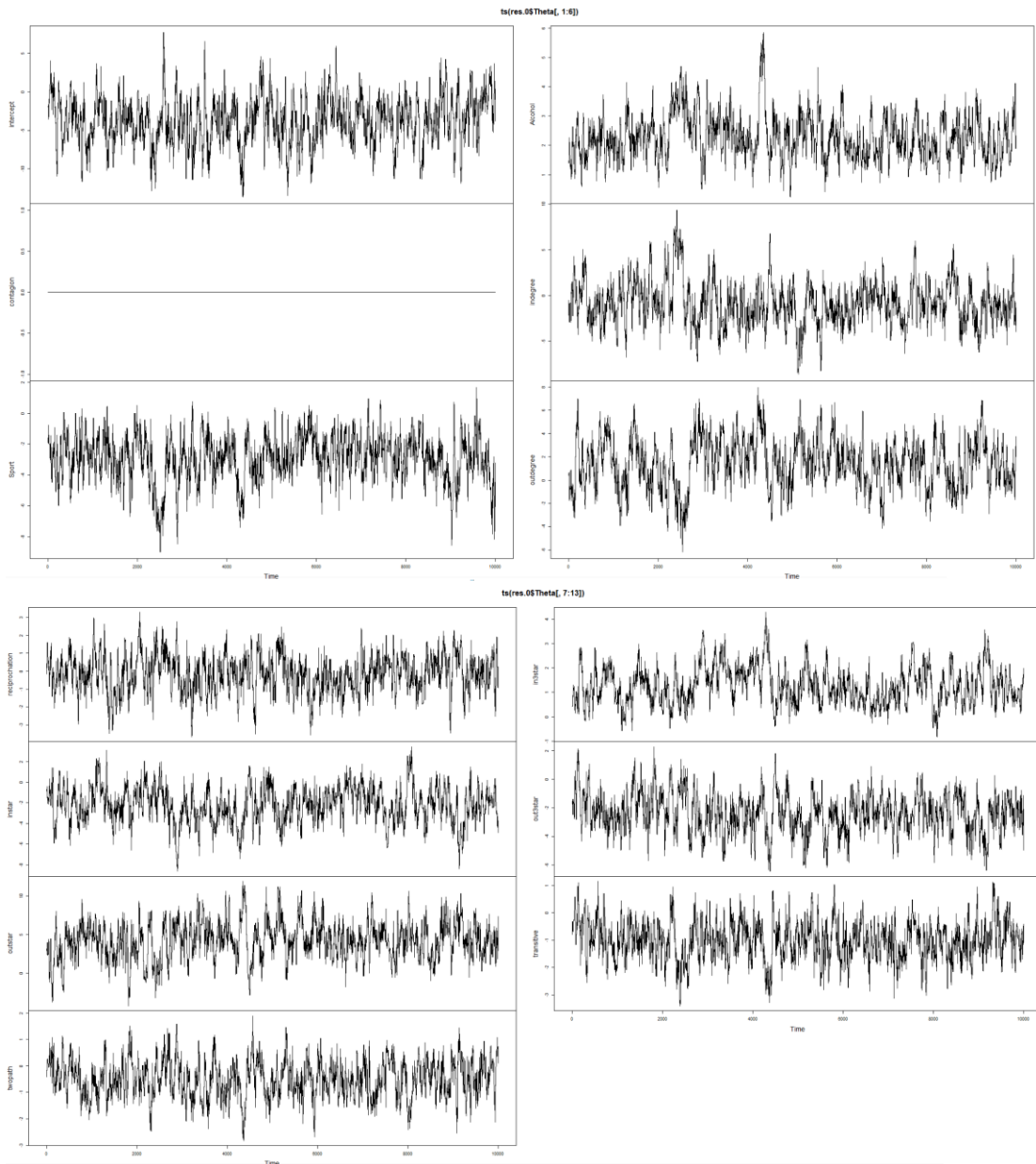
Answer:

Social Network Lab 3: Autologistic Actor Attribute Model (ALAAM)

Student: Yijian Li   Net ID: YLS9426

```
summaries of the posterior draws:
                   mean          sd         ESS      SACF 10      SACF 30
intercept     -3.7290554   3.1462251 148.6946608   0.7140035   0.3656031
contagion      0.0000000   0.0000000   0.0000000
Sport         -2.9093948   1.5268378 158.5460347   0.6923220   0.3978095
Alcohol        2.3124755   0.7735869 125.4719318   0.7691562   0.4774202
indegree      -0.9145913   2.4295910 124.7253118   0.7608525   0.4802950
outdegree      1.3998477   2.1046880 114.4930671   0.7840416   0.5260862
reciprochation -0.1509681  1.0644922 165.1240317   0.6825455   0.2886128
instar        -2.1795562   1.7733316 121.0208219   0.7588592   0.4688075
outstar        4.3190798   2.4304965 147.6602199   0.7347758   0.4291579
twopath       -0.5000923   0.7092356 163.6302055   0.6989230   0.3493415
in3star        1.2740770   0.7742358  89.1923466   0.8172393   0.5684382
out3star      -2.4301748   1.3685556 153.6424004   0.7200068   0.4117515
transitive    -1.0288171   0.7015816 195.4777794   0.6544728   0.2824910
```

Now we can see all the graphs looks better than before, all values are normal distributed around every attribute's mean values.

In every plot, the fluctuations have become less and more stable than it used to be.

Also, the SACF 10 values are mostly below 0.70, SACF 30 values are mostly below 0.30

Basically, we could observe from the MCMC samples:

1. **Intercept** randomly fluctuates around the mean value(-3.72). The graph is generally fit a normal distribution centered around -3.72.
   Ess value is **148(greater than 100),** which means it is a **good value** in our model.
2. **Sport** randomly fluctuates around the mean value(-2.91). The graph is generally fit a normal distribution centered around -2.91.
   Ess value is **158(greater than 100)**, which means it is a **good value** in our model.
3. **Alcohol** randomly fluctuates around the mean value(2.31). The graph is generally fit a normal distribution centered around 2.31.
   Ess value is **125(greater than 100)**, which means it is a **good value** in our model.
4. **indegree** randomly fluctuates around the mean value(-0.91). The graph is generally fit a normal distribution centered around -0.91.
   Ess value is **124(greater than 100)**, which means it is a **good value** in our model.
5. **outdegree** randomly fluctuates around the mean value(1.39). The graph is generally fit a normal distribution centered around 1.39.
   Ess value is **114(greater than 100)**, which means it is a **good value** in our model
6. **reciprochation** randomly fluctuates around the mean value(-0.15). The graph is generally fit a normal distribution centered around -0.15.
   Ess value is **165(greater than 100)**, which means it is a **good value** in our model.
7. **instar** randomly fluctuates around the mean value(-2.17). The graph is generally fit a normal distribution centered around -2.17.
   Ess value is **121(greater than 100)**, which means it is a **good value** in our model.
8. **outstar** randomly fluctuates around the mean value(4.31). The graph is generally fit a normal distribution centered around 4.31.
   Ess value is **147(greater than 100)**, which means it is a **good value** in our model.
9. **twopath** randomly fluctuates around the mean value(-0.50). The graph is generally fit a normal distribution centered around -0.50.
   Ess value is **163(greater than 100)**, which means it is a **good value** in our model.
10. **in3star** randomly fluctuates around the mean value(1.27). The graph is generally fit a normal distribution centered around 1.27.
    Ess value is **89(less than 100)**, which means it is **not a good value** in our model.
11. **out3star** randomly fluctuates around the mean value(-2.43). The graph is generally fit a normal distribution centered around -2.43.
    Ess value is **153(greater than 100)**, which means it is a **good value** in our model.

12. **transitive** randomly fluctuates around the mean value(1.02). The graph is generally fit a normal distribution centered around 1.02.
Ess value is **195(greater than 100)**, which means it is a **good value** in our model.

5. **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.)**

Answer:

	parameter	mean	sd	.025	0.975
1	intercept	-3.709	3.137	-10.228	2.316
2	contagion	0.000	0.000	0.000	0.000
3	Sport	-2.893	1.502	-6.352	-0.291
4	Alcohol	2.281	0.762	1.000	3.907
5	indegree	-0.895	2.389	-5.228	4.754
6	outdegree	1.423	2.107	-2.819	5.410
7	reciprochation	-0.149	1.041	-2.278	1.811
8	instar	-2.177	1.754	-5.893	1.090
9	outstar	4.241	2.442	-0.706	9.178
10	twopath	-0.493	0.705	-1.852	0.863
11	in3star	1.275	0.768	-0.003	2.891
12	out3star	-2.389	1.365	-5.222	0.255
13	transitive	-1.006	0.709	-2.433	0.402

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

DV: smoke regularly, IV: alcohol regularly

Alcohol:

- The 95% confidence interval is from 1.00 to 3.907 (**not cross 0**)
- Estimate is 2.281, which is **positive**.

So, there is a **positive** relationship between using alcohol and smoking and the **hypothesis 1 is true**.

- Exp(2.281) = **9.786462**.

Teen girls who drink alcohol are **9.79 time more likely** to smoke than those who do not drink alcohol.

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

DV: smoke regularly, IV: are friends more

Indegree:

- The 95% confidence interval is from -5.228 to 4.754 (**crossing 0**)

The hypothesis 2 is not correct.

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

DV: smoke regularly, IV: have more friends

Outdegree:

- The 95% confidence interval is from -2.819 to 5.410 (**crossing 0**)

The hypothesis 3 is not correct.

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

DV: smoke regularly, IV: have reciprocal friends

Reciprochation:

- The 95% confidence interval is from -2.278 to 1.811 (**crossing 0**)

The hypothesis 4 is not correct.

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

DV: smoke regularly, IV: are friends by many people

Instar:

- The 95% confidence interval is from -5.893 to 1.090 (**crossing 0**)

The hypothesis 5 is not correct.

B. Model 2 (simple social contagion model):

1. **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.**

Answer:

```
summaries of the posterior draws:
              mean          sd        ESS       SACF 10      SACF 30
intercept -2.10672575  1.68590366 19.12547072  0.61233019  0.24549250
contagion  0.64260367  0.27628137 26.83801918  0.48539991  0.04646922
Sport     -0.82722328  0.61766736 35.88180834  0.50945733  0.10076034
Alcohol    0.86304672  0.36390863 23.09480168  0.55270777  0.21140131
outdegree -0.42737506  0.29940313 19.21602738  0.57941285  0.03630049
```

2. **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?)**

Answer:
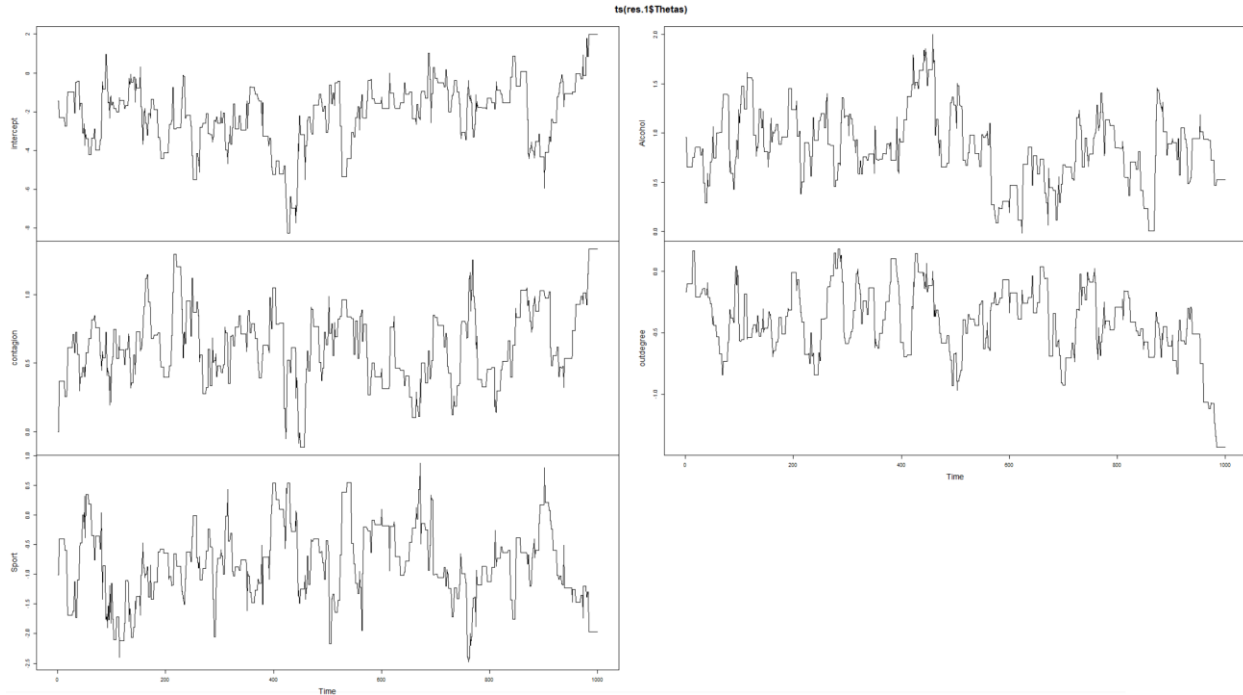
1. The graphs of **intercept**, have a little bit more fluctuations than others, but still fluctuates around its mean value(-2.106)

2. For **contagion**, it randomly fluctuates around the mean value(0.64). The graph is generally fit a normal distribution centered around 0.64.

3. The graphs of **sport**, have a little bit more fluctuations than others, but still fluctuates around its mean value(-0.827)

4. The graphs of **alcohol**, it randomly fluctuates around the mean value(0.86). The graph is generally fit a normal distribution centered around 0.86.

5. The graphs of **outdegree**, have a little bit more fluctuations at the end than others, but still fluctuates around its mean value(-0.427)

3. **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.**

Answer:

```
summaries of the posterior draws:
              mean          sd         ESS     SACF 10     SACF 30
intercept  -1.4614059   1.9903944  90.6251170  0.6858574   0.4308366
contagion   0.6732435   0.2844370 127.6358663  0.5749448   0.2422926
Sport      -1.0914333   0.9007501  73.2681104  0.7460368   0.4316585
Alcohol     0.8882109   0.4084487 103.4779447  0.6310611   0.2622146
outdegree  -0.5713984   0.3109967 125.8873008  0.5616859   0.2096713
```

4. **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?**

Answer:
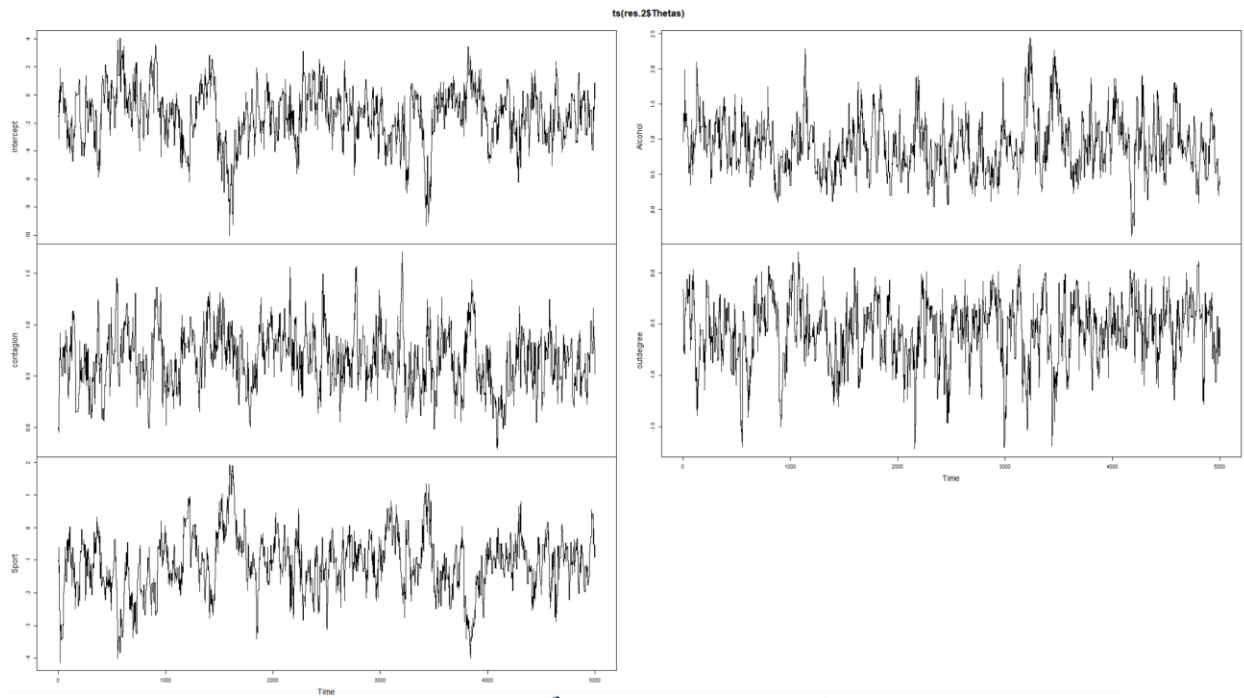
Now we can see all the graphs looks better than before, all values are normal distributed around every attribute's mean values.

In every plot, the fluctuations have become less and more stable than it used to be.

Also, the SACF 10 values are mostly below 0.70, SACF 30 values are mostly below 0.30

1. **Intercept** randomly fluctuates around the mean value(-1.46). The graph is generally fit a normal distribution centered around -1.46.
2. **contagion** randomly fluctuates around the mean value(0.67). The graph is generally fit a normal distribution centered around 0.67.
3. **Alcohol** randomly fluctuates around the mean value(0.88). The graph is generally fit a normal distribution centered around 0.88.
4. **Sport** randomly fluctuates around the mean value(-1.09). The graph is generally fit a normal distribution centered around –1.09.

5. **Outdegree** randomly fluctuates around the mean value(-0.57). The graph is generally fit a normal distribution centered around -0.57.

5. 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.)

Answer:

	parameter	mean	sd	.025	0.975
1	intercept	-1.462	1.973	-5.852	1.924
2	contagion	0.663	0.286	0.124	1.252
3	Sport	-1.117	0.895	-3.027	0.728
4	Alcohol	0.906	0.403	0.204	1.810
5	outdegree	-0.571	0.308	-1.251	-0.051

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

DV: smoke regularly, IV: others smoke regularly

contagion:

- The 95% confidence interval is from 0.124 to 1.252 (**not cross 0**)
- Estimate is 0.663, which is **positive**.

So, there is a **positive** relationship between one person smoke regularly and others smoke regularly, the **hypothesis 6 is true**.

- Exp(0.663) = **1.940605**.

Teen girls who are affected to smoke are **1.94 time more likely** to smoke than those who do not have smokers around them.

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

DV: smoke regularly, IV: exercise regularly

sport:

- The 95% confidence interval is from -3.027 to 0.728 (**crossing 0**)

The hypothesis 7 is not correct.

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

DV: smoke regularly, IV: drink alcohol regularly

contagion:

- The 95% confidence interval is from 0.204 to 1.810 (**not cross 0**)
- Estimate is 0.906, which is **positive**.

So, there is a **positive** relationship between using alcohol and smoking and the **hypothesis 8 is true**.

- Exp(0.906) = **2.474405**.

Teen girls who drink alcohol are **2.47time more likely** to smoke than those who do not drink alcohol.

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

DV: smoke regularly, IV: more friends

outdegree:

- The 95% confidence interval is from -1.251 to -0.051 (**not cross 0**)
- Estimate is -0.571, which is **negative**.

So, there is a **negative** relationship between one person smoke regularly and have more friends, the **hypothesis 9 is false**.

- Exp(-0.571) = **0.5649602**.

Teen girls who have more friends are **0.56 time less likely** to smoke than those who have fewer friends.

Part V: Goodness-of-fit test

1. **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).**

Answer:

Social Network Lab 3: Autologistic Actor Attribute Model (ALAAM)

Student: Yijian Li   Net ID: YLS9426

```
                      obs     mean    p-val
intercept          17.000  17.794   0.201
simple cont.       25.000  27.144   0.223
recip cont.         7.000   9.476   0.133
indirect cont.     59.000  47.886   0.148
closedind cont.    22.000  23.566   0.228
transitive cont.   13.000  18.066   0.154
indegree           43.000  43.580   0.235
outdegree          40.000  45.470   0.167
twopaths           26.000  29.902   0.162
out2star           57.000  62.952   0.166
in2star            44.000  42.238   0.218
outtria            88.000  95.522   0.209
intria             71.000  67.086   0.321
transtri           22.000  20.464   0.206
indirecct ties     34.000  35.458   0.233
```

As we could see from the chart, all the p-vals we calculated are **all larger than 0.05**, which means they are **good,** and **all the attributes are good fit**.

The **model 1 is a good model for observation and analysis here.**