IEMS 441: Social Network Analysis
Tyler Maule
November 13th, 2021

Lab 3a: Emergence of Network Dynamics—Stochastic Actor Oriented Models

PART I: Constructing Hypotheses (30 points)

1. Low Outdegree Density/ outdegree

Hypothesis 1: The probability of having friendship relation between students will be lower over time than expected by random chance.

Relational Hypotheses:

2. Reciprocity/reciprocity **(5 points)**

Hypothesis 2: The probability of reciprocity (that is, the probability that A mentions B as a friend given that B mentions A as a friend) will be higher over time than expected by random chance.

3. Transitivity/gwespFF **(5 points)**

Hypothesis 3: The probability of transitivity (that is, the probability that A and B are connected through friendship given that A is connected to C and B is connected to C) will be higher over time than expected by random chance.

4. Ego's drug behavior/ egoX **(5 points)**

Hypothesis 4: The probability of having friendship relations given the ego's drug behavior (behavior of the node in question) will be lower over time than expected by random chance.

5. Alter's drug behavior/ altX **(5 points)**

Hypothesis 5: The probability of having friendship relations given alters' drug behavior (behavior of the node in question's neighbors) will be lower over time than expected by random chance.

6. Homophily on the basis of drug behavior/sameX **(5 points)**

Hypothesis 6: The probability of having friendship relations given that two nodes have the same drug behavior will be higher over time than expected by random chance.

Drug Behavior Hypothesis:

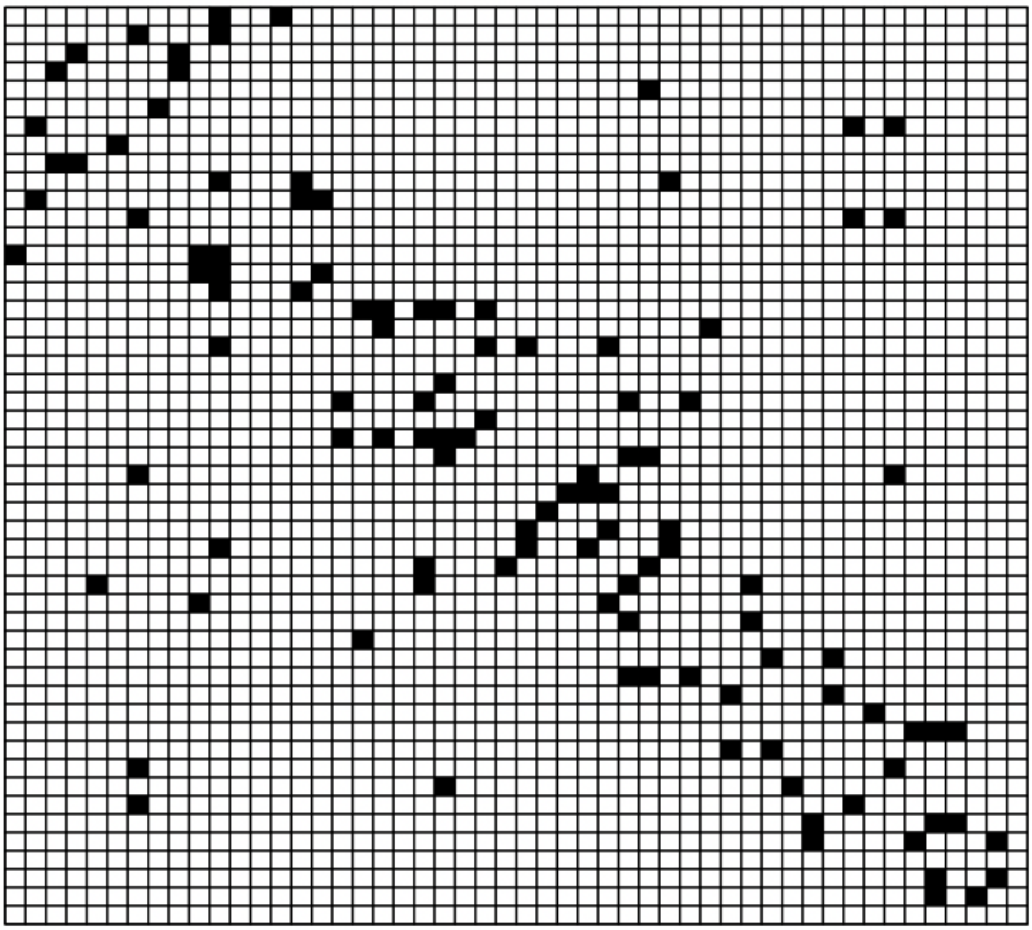
7. Assimilation of drug behavior/totSim **(5 points)**

Hypothesis 7: The probability of drug behavior spreading through the friendship network will be higher over time than expected by random chance.

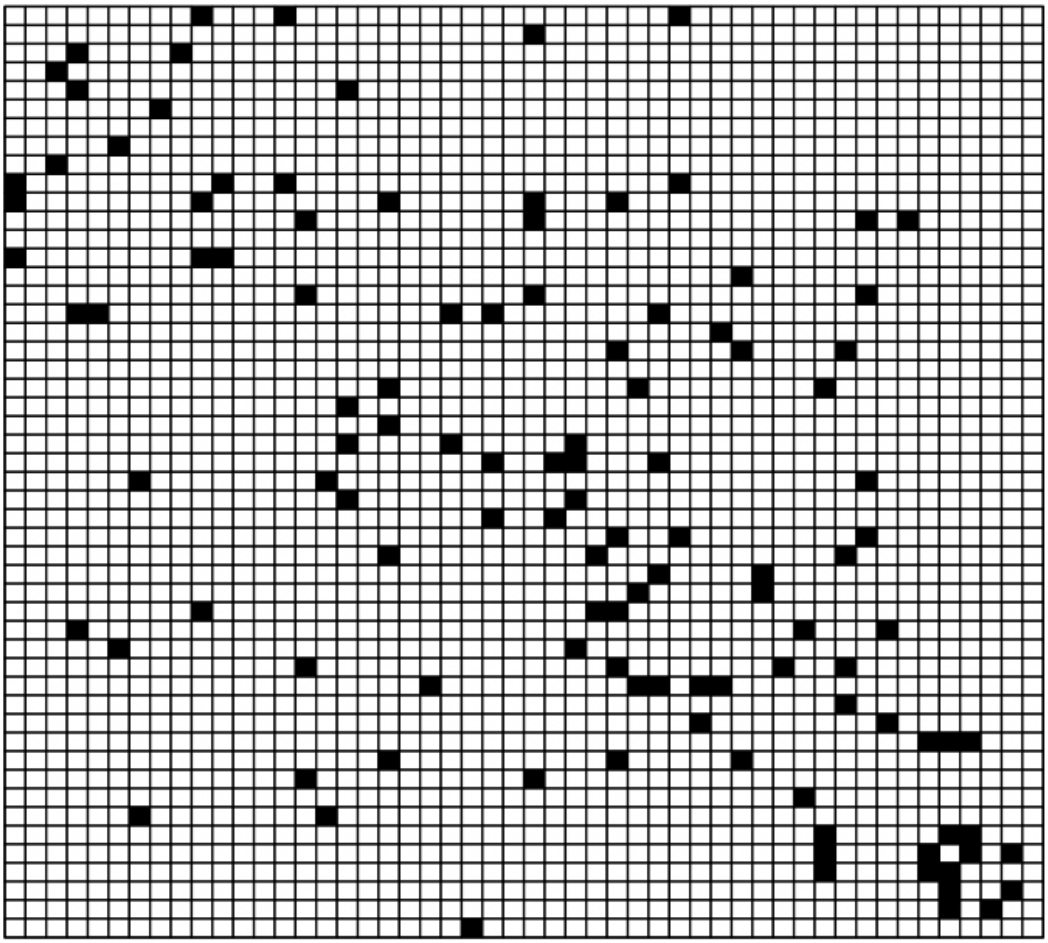
PART II: Hypothesis Testing (70 points)

1. A visual inspection of the adjacency matrices may help in highlighting how friendship changes at the three time. Include the sociomatrix plots in your report. Discuss what you observe from the plots (e.g., How does friendship change over time? Are the plots becoming denser over time? Is friendship between students mutual? Is there anyone who is nominated a lot by others? Is there anyone who nominates a lot of friends?) **(10 points)**

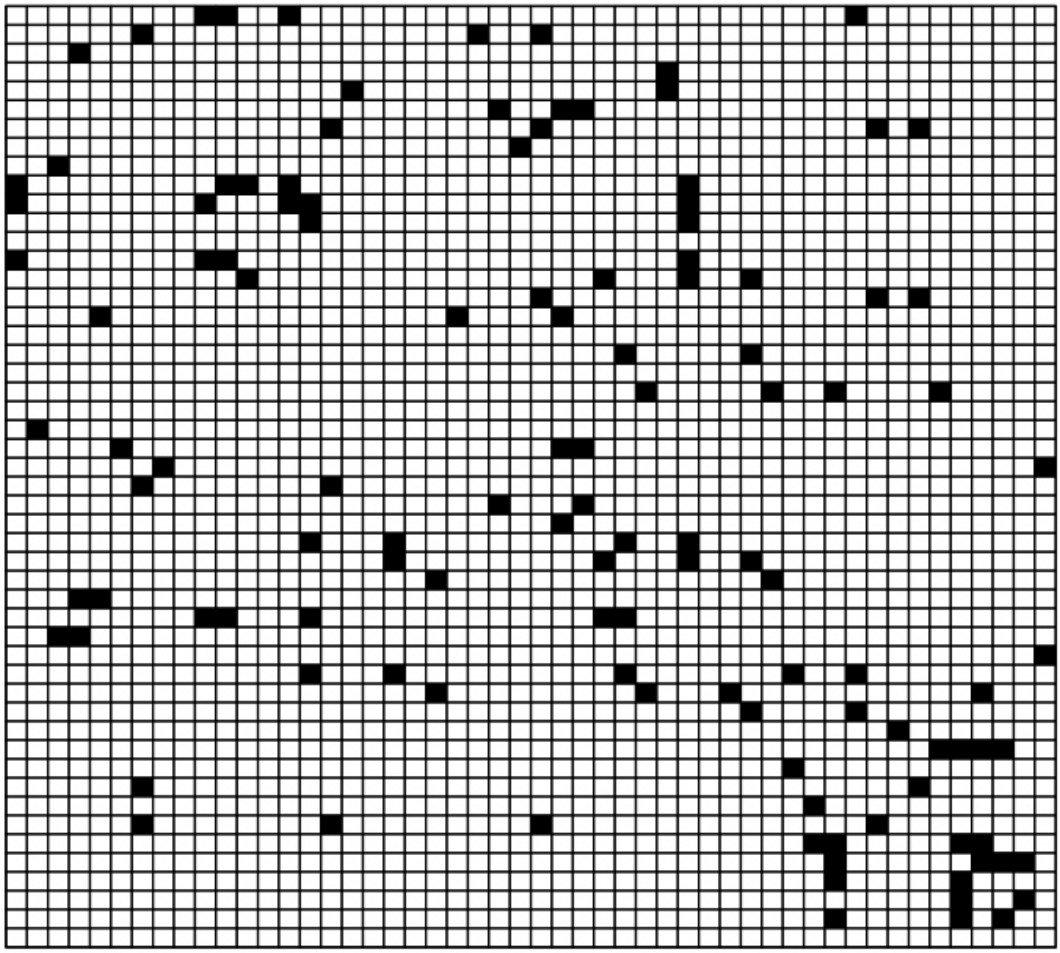
Note below the sociomatrix at time *t=1*. At a glance, it seems that ties in this network are largely symmetric; the filled in cells are close to a mirror image across the matrix diagonal, indicating that for most nodes *A*, *B* if *A* indicates that *B* is their friend *B* will indicate that *A* is their friend. The network also appears relatively sparse, as the majority of actors are nominated by three or fewer friends (with some exceptions being named by up to seven, which have a higher indegree—nominating more friends). Outdegree appears even lower generally, with a maximum of five friends nominated by two actors (most nominate around three friends).



Now examine the second sociomatrix below, at *t=2*. If anything, the updated sociomatrix appears marginally more sparse than the initial matrix. The maximum outdegree appears the same, at five, but the average outdegree may have decreased a bit. Changes in indegree are clearer, with an apparent new maximum of five and a slightly decreased average indegree. Directional ties have moved from the diagonal of the matrix outwards, demonstrating clear changes in friendship ties. While it's more difficult to tell from this matrix, it seems that most ties are mutual (similar to the first sociomatrix).



Finally, take a look at the following sociomatrix that represents *t=3*. Again, the matrix appears different than its prior iteration, although the difference between *t=2,3* appears lesser than the difference between *t=1,2*. We see some consistent groups of friends that are (mostly) mutual, such as the cluster of friends in the lower right-hand corner of the matrix. Overall, the matrix again appears more sparse than its predecessor. On the whole, friendship ties still appear similarly mutual in nature. Maximum outdegree is still five, and there don't appear to be global changes to the average outdegree. In addition, the change in indegree is slight; if anything, actors' average number of incoming nominations is a little higher.



In sum, we're seeing a friendship matrix gently increase in sparsity and reciprocity over time. Average outdegree decreases modestly over time, and indegree fluctuates slightly. While there are changes in the friendship ties, the change between the first two sociomatrices is noticeably larger than the difference between the later two sociomatrices.

Since these differences are so slight from my perspective, I feel strongly that quantitative analysis is needed to affirm or dispute my thinking.

2. Create a *siena* data object including the longitudinal friendship networks and the drug behavioral variable. Then run print01report function which creates an output file in your working directory. Using your text editor, open the output file (if you use the provided

script, s50_3_init.out) where you can see data descriptions. In the output file, how many friendship relations were created and dissolved between period 1 and 2? How many students increased their use of drug or decreased the use of drug between the same periods? **(4 points)**

Between periods 1 and 2, 59 friendship relations were created, 56 were dissolved, and 57 were retained. During the same period, 2 actors decreased their drug use, 5 increased their drug use, and 43 saw no change in their drug use.

The timespan between periods 2 and 3 saw 56 new friendship relations, the dissolution of 50 friendship relations, and the maintenance of 66 friendship relations. Over that same timespan, 3 actors decreased their drug use, 13 increased drug use, and 34 had no change in their drug use.

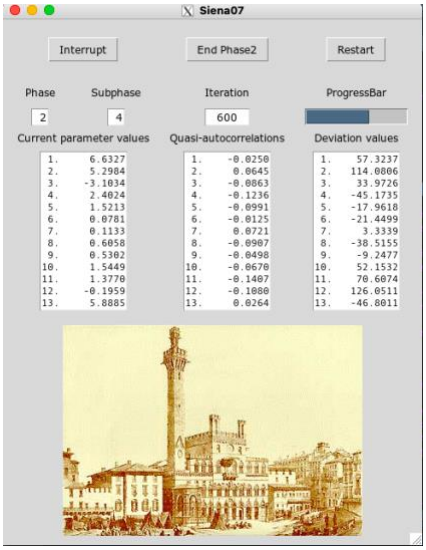
3. Using your hypotheses, you can begin to construct a list of parameters (effects) to test using your Siena model. Create a data frame of effects using the getEffects function. The created data frame will include a number of extra properties for use with RSiena. Include the effects of triadic closure, the effects of drug use on friendship formation (i.e. the effects of the ego drug behavior, the alter drug behavior, and if ego and alter are the same in their drug behavior –both drug user or non-drug user).

Done—see an excerpt of the table below.

row	name	effectName	shortName	type	inter1	inter2	parm	interactionType
1	friendship	constant friendship rate (period 1)	Rate	rate			0	
2	friendship	constant friendship rate (period 2)	Rate	rate			0	
3	friendship	outdegree effect on rate friendship	outRate	rate			0	
4	friendship	indegree effect on rate friendship	inRate	rate			0	
5	friendship	reciprocity effect on rate friendship	recipRate	rate			0	
6	friendship	effect 1/outdegree on rate friendship	outRateInv	rate			0	
7	friendship	effect ln(outdegree+1) on rate friendship	outRateLog	rate			1	
8	friendship	effect drugbeh on rate	RateX	rate	drugbeh		0	

4. Include the effects of all of a node's friends' drug behavior on the node's own drug behavior, the specified model and effects to the data using the function siena07. A new window labeled "Siena07" with a picture of an old building should pop up, showing the iterations of simulations R goes through. The window should close after the simulations complete.

Done—see the screenshot below.



5. Type ans1 to view your results.

(a) Include a table including convergence t-ratios and overall maximum convergence ratio in your report.

Find below the table produced by R, including convergence t-ratios and the overall maximum convergence ratio.

```
> ans1
Estimates, standard errors and convergence t-ratios

                                          Estimate   Standard   Convergence
                                                       Error      t-ratio
Network Dynamics
   1. rate constant friendship rate (period 1)  6.6870  ( 1.1659   )   -0.0207
   2. rate constant friendship rate (period 2)  5.3658  ( 0.9801   )    0.0335
   3. eval outdegree (density)                 -3.1016  ( 0.2211   )   -0.0018
   4. eval reciprocity                          2.4293  ( 0.1779   )   -0.0047
   5. eval GWESP I -> K -> J (69)               1.5186  ( 0.1721   )   -0.0157
   6. eval drugbeh alter                        0.0739  ( 0.1262   )    0.0415
   7. eval drugbeh ego                          0.1167  ( 0.1385   )    0.0158
   8. eval same drugbeh                         0.5876  ( 0.2517   )   -0.0203

Behavior Dynamics
   9. rate rate drugbeh (period 1)              0.5335  ( 0.2516   )    0.0051
  10. rate rate drugbeh (period 2)              1.5772  ( 0.6255   )   -0.0176
  11. eval drugbeh linear shape                 1.4828  ( 1.7475   )   -0.0196
  12. eval drugbeh quadratic shape             -0.1637  ( 0.4725   )    0.0148
  13. eval drugbeh total similarity             5.9694  ( 5.7302   )    0.0045

Overall maximum convergence ratio:    0.0906


Total of 2963 iteration steps.
```

(b) Explain whether your variables and model are converged based on your convergence t-ratios and overall maximum convergence ratio. See the hint (*).

*[Remember, a |t-ratio| < 0.1 means that the parameter converged to a reliable value. An overall maximum convergence ratio < 0.25 indicates the model convergence. A parameter is significant if the $\frac{|estimate|}{standard\ error} > 1.96$ (i.e., $p < 0.05$)]. **(46 points)**

The convergence t-ratio with the *maximum* absolute value among the ratios corresponding to our model parameters has a value of 0.0415. As this value still falls below the 0.1 threshold, we can assert that all our variables converged to reliable values.

Our overall maximum convergence ratio is 0.0906, which is clearly lower than the threshold of 0.25. This result provides evidence that the model converged overall.

Therefore the model and all variables converged to an acceptable degree.

(c) Include another table including the estimates, standard errors and p-values (or estimate/standard error) in your report.

The relevant table, reproduced below, contains estimates, standard errors, and descriptors of binned p-values.

	Model 1
constant friendship rate (period 1)	6.69^{***}
	(1.17)
constant friendship rate (period 2)	5.37^{***}
	(0.98)
outdegree (density)	-3.10^{***}
	(0.22)
reciprocity	2.43^{***}
	(0.18)
GWESP I -> K -> J (69)	1.52^{***}
	(0.17)
drugbeh alter	0.07
	(0.13)
drugbeh ego	0.12
	(0.14)
same drugbeh	0.59^{*}
	(0.25)
rate drugbeh (period 1)	0.53^{*}
	(0.25)
rate drugbeh (period 2)	1.58^{*}
	(0.63)
drugbeh linear shape	1.48
	(1.75)
drugbeh quadratic shape	−0.16
	(0.47)
drugbeh total similarity	5.97
	(5.73)
Iterations	2963

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 1: Statistical models

(d) Use the estimates and p-value (or estimate/standard error) to explain whether your hypotheses (7 hypotheses in total) are supported or not. Provide interpretations of the estimates and discuss if the results make sense. **When you interpret the results, you should convert log-odds ratios (estimates) into either log-odds or probabilities.**

<u>1. Low Outdegree Density/ outdegree</u>

Hypothesis 1: The probability of having friendship relation between students will be lower over time than expected by random chance.

Parameter *outdegree (density)* has an estimated value of -3.10, which can be exponentiated to find 0.0450492. The associated p-value is below 0.001, which means that at the $\alpha = 0.05$ level we can reject the null hypothesis that there is no change in *outdegree (density)* over time.

Based upon this estimate, the odds of individuals reporting a friendship tie is 0.04 times lower over time relative to the odds that individuals do not report a friendship tie. These results support our hypothesis that the probability of reported friendship ties will be lower over time than expected by random chance, to a statistically significant degree.

This finding makes sense, as most teenagers tend to have small circles of friends and may be hesitant to attach their identity or reputation to that of their peers. As a result there would be a low tendency to nominate peers as friends, and thus a low outdegree density.

Relational Hypotheses:

2. Reciprocity/reciprocity

Hypothesis 2: The probability of reciprocity (that is, the probability that A mentions B as a friend given that B mentions A as a friend) will be higher over time than expected by random chance.

Parameter *reciprocity* has an estimated value of 2.43, which can be exponentiated to find 11.35888. The associated p-value is below 0.001, which means that at the $\alpha = 0.05$ level we can reject the null hypothesis that there is no change in *reciprocity* over time.

Based upon this estimate, the odds of individuals reporting reciprocal friendship ties is 11.35888 times lower over time relative to the odds that individuals do not share a reciprocal friendship tie. These results support our hypothesis that the probability of reported reciprocal friendship ties will be higher over time than expected by random chance, to a statistically significant degree.

3. Transitivity/gwespFF

Hypothesis 3: The probability of transitivity (that is, the probability that A and B are connected through friendship given that A is connected to C and B is connected to C) will be higher over time than expected by random chance.

Parameter *GWESP I->K->J* has an estimated value of 1.52, which can be exponentiated to find 4.572225. The associated p-value is below 0.001, which means that at the $\alpha = 0.05$ level we can reject the null hypothesis that there is no change in *GWESP I->K->J* over time.

Based upon this estimate, the odds of transitivity in friendship ties is 4.572225 times lower over time relative to the odds that transitivity in friendship ties does not occur. These results support our hypothesis that the probability of transitivity in friendship ties will be higher over time than expected by random chance, to a statistically significant degree.

4. Ego's drug behavior/ egoX

Hypothesis 4: The probability of having friendship relations given the ego's drug behavior (behavior of the node in question) will be lower over time than expected by random chance.

Parameter *drugbeh ego* has an estimated value of 0.12, which can be exponentiated to find 1.127497. The associated p-value is above or equal to 0.05, which means that at the $\alpha = 0.05$ level we *cannot* reject the null hypothesis that there is no change in *drugbeh ego* over time.

Based upon this estimate, the odds of friendship ties being reported given the ego's drug behavior is 1.127497 times lower over time relative to the odds that friendship ties do not occur given the ego's drug behavior. These results support our hypothesis that the probability of friendship ties given the ego's drug behavior will be lower over time than expected by random chance, but *not* to a statistically significant degree.

This finding makes sense, as the drug in question is marijuana (for "harder" drugs my prediction would be different). While I'd think that homophily would predict ties best, I don't believe that someone would be likely to nominate a greater or lesser number of friends depending on their own amount of drug use. If anything, I'd guess that individuals who use more might have a tighter circle of friends with whom they use or exchange drugs, and thus would nominate a marginally smaller number of friends.

5. Alter's drug behavior/ altX

Hypothesis 5: The probability of having friendship relations given alters' drug behavior (behavior of the node in question's neighbors) will be lower over time than expected by random chance.

Parameter *drugbeh alter* has an estimated value of 0.07, which can be exponentiated to find 1.072508. The associated p-value is above or equal to 0.05, which means that at the $\alpha = 0.05$ level we *cannot* reject the null hypothesis that there is no change in *drugbeh alter* over time.

Based upon this estimate, the odds of friendship ties being reported given the alter's drug behavior is 1.072508 times lower over time relative to the odds that friendship ties do not occur given the alter's drug behavior. These results contradict our hypothesis that the probability of reported friendship ties given the alters' drug behavior will be lower over time than expected by random chance, but *not* to a statistically significant degree.

Since teens with any level of drug use are likely to find a community of some size that has similar drug habits, I don't think peers' drug use habits would affect the number of alters that would nominate them as a friend (although the group of alters itself may change). So this finding makes sense to me, for reasons similar to those given for hypothesis 4.

6. Homophily on the basis of drug behavior/sameX

Hypothesis 6: The probability of having friendship relations given that two nodes have the same drug behavior will be higher over time than expected by random chance.

Parameter *same drugbeh* has an estimated value of 0.59, which can be exponentiated to find 1.803988. The associated p-value is below 0.05, which means that at the $\alpha = 0.05$ level we can reject the null hypothesis that there is no change in *same drugbeh* over time.

Based upon this estimate, the odds of friendship ties being reported given that the ego and alter share the same drug behavior is 1.803988 times lower over time relative to the odds that friendship ties do not occur given that the ego and alter have the same drug behavior. These results support our hypothesis that the probability of reported friendship ties between teens with similar levels of drug use will be higher over time than expected by random chance, to a statistically significant degree.

This conclusion isn't surprising to me. Since marijuana use is often a social activity, and at the least requires social ties to acquire, I'd think that users would be friends with other users. By the same token, teens who didn't want to engage in drug use may be hesitant to spend time with others who use to a greater degree.

Drug Behavior Hypothesis:

7. Assimilation of drug behavior/totSim

Hypothesis 7: The probability of drug behavior spreading through the friendship network will be higher over time than expected by random chance.

Parameter *drugbeh total similarity* has an estimated value of 5.97, which can be exponentiated to find 391.5057. The associated p-value is above or equal to 0.05, which means that at the $\alpha = 0.05$ level we *cannot* reject the null hypothesis that there is no change in *drugbeh total similarity* over time.

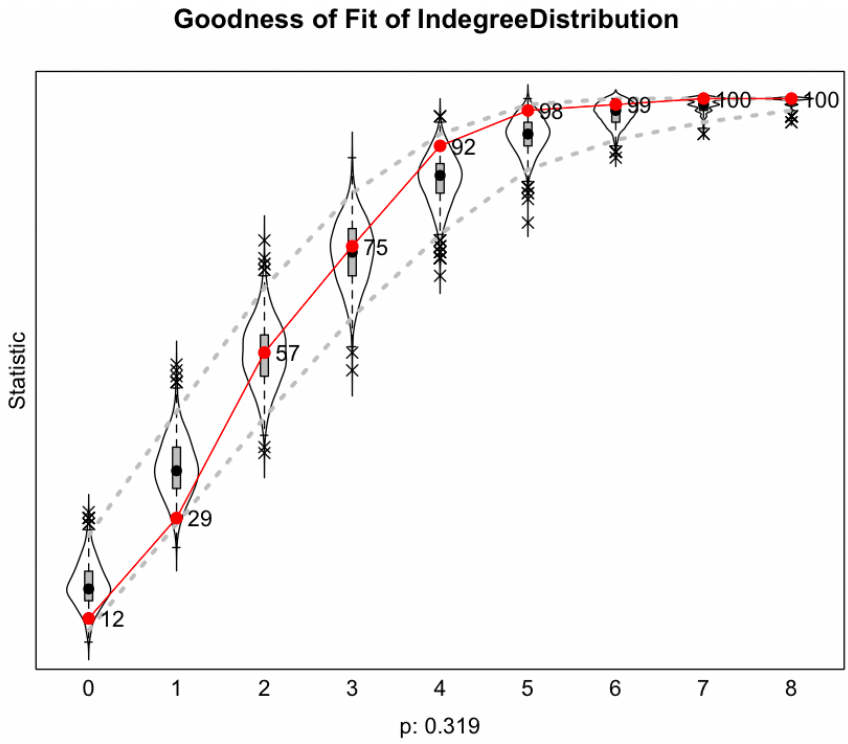
Based upon this estimate, the odds of drug behavior spreading through the friendship network is 391.5057 times lower over time relative to the odds that drug behavior does not diffuse across the network. These results support our hypothesis that the probability of reported friendship ties between peers with similar levels of drug use will be higher over time than expected by random chance, but not to a statistically significant degree.

This finding also seems reasonable to me. I do believe that teens would increase their use of marijuana as they grow older, become more independent, and come into contact with a greater number of other drug users. But the 13-16 range is a little younger than I'd expect drug use to spread in the network to a statistically significant degree.

6. Report the goodness of fit for your model with regard to in-degree and out-degree distributions. Include the plots and interpret the results of each plot. **(10 points)**

See the goodness of fit indegree distribution plot below.

As the associated p-value of 0.316 lies above 0.05, we find that our model is a good fit to the empirical friendship network in terms of indegree. The 95% confidence interval for the estimated indegree parameter statistics all contain the true empirical statistic.



See the plot corresponding to outdegree below.

Since the associated p-value of 0.295 exceeds the 0.05 threshold, we know that our model does a good job of fitting the true friendship network with respect to outdegree. We can verify this visually, as the real network's outdegree statistics all fall within the 95% confidence intervals based on the simulated model networks.