**Social Network Lab 2: Exponential Random Graph Models**

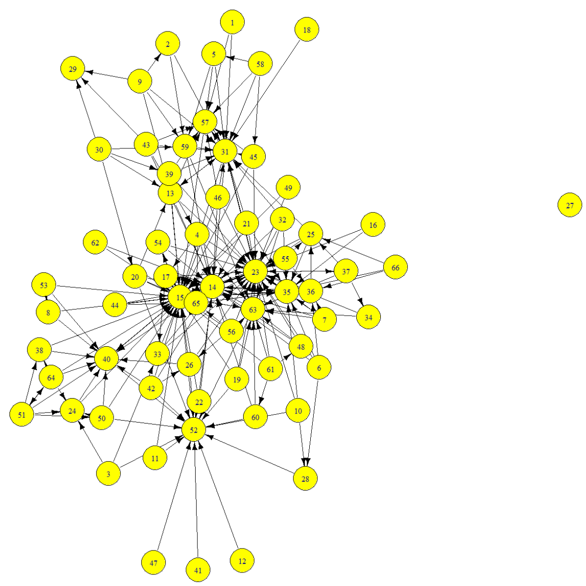
*Student: Yijian Li   Net ID: YLS9426*

# Part I: Building and Visualizing the Networks

1. **Plot the base (Advice) network and include it in your report. Explain whether this plot seems, at a glance, to match what you would expect to see if hypothesis 1 were true. Think of this as just a basic descriptive check – we will perform a more rigorous statistical test in part II of the lab.**

Answer:

*Hypothesis 1: There will be indegree popularity effects – That is, a tendency for a small number of employees to be sought out for advice from many others (as opposed to advice seeking behaviors being spread evenly amongst all employees).*



As we can see, there are nodes(employees) 40, 52, 23, 14, 63, and some other nodes are be sought out for advice from many others instead of finding advisors evenly in the network.

And we could roughly conclude that the hypothesis 1 is true for our network.

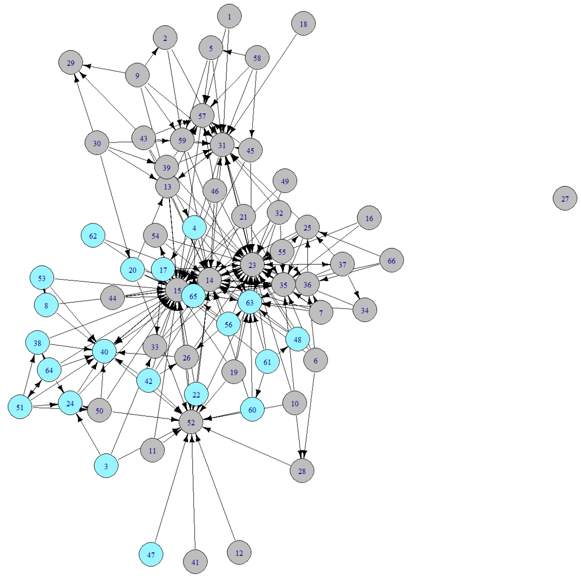
2. **Plot the base network with the nodes now colored based on sex and include it in your report. Explain whether this plot seems to match what you would expect to see if hypothesis 3 were true.**

Answer:

*Hypothesis 3: There will be homophily based upon the sex of individuals, in terms of who employees go to for advice.*

**Social Network Lab 2: Exponential Random Graph Models**

*Student: Yijian Li   Net ID: YLS9426*



In the chart, the blue represents female and gray represents males.

We could clearly see that 51 are seeking for advice with all females and there are lots of blue nodes only seek for advice for blue nodes in the chart.

For node 40, 80% of people who contact to her for advice are females.

Therefore, we can say the hypothesis 3 is true.

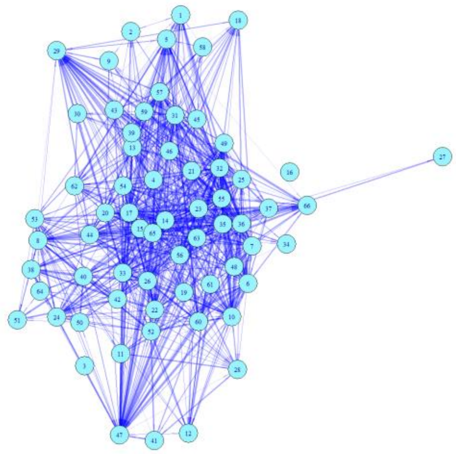
3. **Plot the covariate network and include it in your report. Comparing the network plots, explain whether this plot seems to match what you would expect to see if hypothesis 4 were true.**

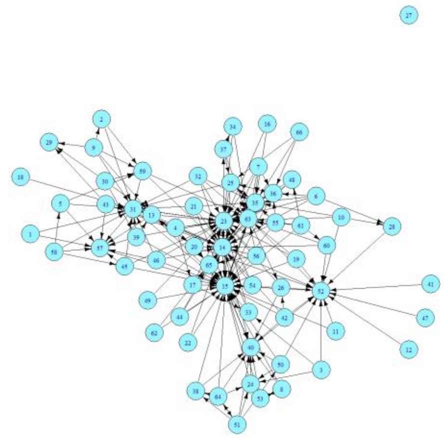
Answer:

*Hypothesis 4: Employees who message someone more frequently on ESM will be more likely to report going to that person for advice.*

**Social Network Lab 2: Exponential Random Graph Models**

*Student: Yijian Li   Net ID: YLS9426*





From the chart, we could find out that nodes 14, 15, 23, 65, 52, 31 all have frequently messages with others, so they also receive more people who seek for advice.

For example, node 3 has frequent messages with node 42 and we also see node 3 seek for advice with node 42 in right chart.

In sum, we could conclude that the hypothesis 4 is true for our graph.

# PART II: Model Estimation

1. **Build two ERGM models to test the hypotheses using the different network statistics described below and include the results (screenshot of the model output tables from the R console) in your report. Fit model 1 (simple model) and model 2 (complex model) using the terms already specified in the R script provided to you.**

Answer:

```
> summary(model1)
Call:
ergm(formula = advice ~ edges + mutual + edgecov(hundreds_messages) +
    nodemix("leader", base = 3), constraints = ~bd(maxout = 5))

Monte Carlo Maximum Likelihood Results:

                          Estimate Std. Error MCMC % z value Pr(>|z|)
edges                     -0.67522    0.13364      0  -5.053  < 1e-04 ***
mutual                     1.12640    0.41419      0   2.720 0.006537 **
edgecov.hundreds_messages  0.33658    0.09489      0   3.547 0.000389 ***
mix.leader.0.0            -2.79079    0.15668      0 -17.813  < 1e-04 ***
mix.leader.1.0            -3.43220    0.48343      0  -7.100  < 1e-04 ***
mix.leader.1.1            -0.10109    0.42993      0  -0.235 0.814104
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

     Null Deviance:    0.0  on 4290  degrees of freedom
 Residual Deviance: -720.6  on 4284  degrees of freedom

Note that the null model likelihood and deviance are defined to be 0. This means that all likelihood-based
inference (LRT, Analysis of Deviance, AIC, BIC, etc.) is only valid between models with the same reference
distribution and constraints.

AIC: -708.6  BIC: -670.4  (Smaller is better. MC Std. Err. = 1.611)
```

# Social Network Lab 2: Exponential Random Graph Models

*Student: Yijian Li   Net ID: YLS9426*

```
> summary(model2)
Call:
ergm(formula = advice ~ mutual + gwidegree(log(2), fixed = T) +
    gwodegree(2, fixed = T, cutoff = 5) + dgwesp(log(2), type = "OTP",
    fixed = T, cutoff = 5) + nodematch("female") + nodemix("leader",
    base = 3) + nodematch("department") + nodeicov("office") +
    nodeocov("office") + diff("tenure") + edgecov(hundreds_messages),
    constraints = ~bd(maxout = 5), control = control.ergm(MCMC.effectiveSize = 50))

Monte Carlo Maximum Likelihood Results:

                                  Estimate Std. Error MCMC % z value Pr(>|z|)
mutual                            -0.44041    0.50143      0  -0.878  0.37977
gwideg.fixed.0.693147180559945    -2.42898    0.33796      0  -7.187  < 1e-04 ***
gwodeg.fixed.2                    -3.62608    0.27495      0 -13.188  < 1e-04 ***
gwesp.OTP.fixed.0.693147180559945  0.64096    0.09721      0   6.593  < 1e-04 ***
nodematch.female                   0.18858    0.14273      0   1.321  0.18641
mix.leader.0.0                    -1.10205    0.21414      0  -5.146  < 1e-04 ***
mix.leader.1.0                    -1.54410    0.54739      0  -2.821  0.00479 **
mix.leader.1.1                    -0.46464    0.53376      0  -0.870  0.38403
nodematch.department               2.11496    0.17547      0  12.053  < 1e-04 ***
nodeicov.office                   -0.23915    0.14086      0  -1.698  0.08955 .
nodeocov.office                    0.03340    0.19261      0   0.173  0.86232
diff.t-h.tenure                   -0.14398    0.02399      0  -6.001  < 1e-04 ***
edgecov.hundreds_messages          0.41520    0.09426      0   4.405  < 1e-04 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

     Null Deviance:     0  on 4290  degrees of freedom
 Residual Deviance: -1126  on 4277  degrees of freedom

Note that the null model likelihood and deviance are defined to be 0. This means that all
likelihood-based inference (LRT, Analysis of Deviance, AIC, BIC, etc.) is only valid
between models with the same reference distribution and constraints.

AIC: -1100  BIC: -1018  (Smaller is better. MC Std. Err. = 1.001)
```

# Social Network Lab 2: Exponential Random Graph Models

*Student: Yijian Li   Net ID: YLS9426*

| | model1 | model2 |
|---|---|---|
| edges | -0.68 *** | |
| | (0.13) | |
| mutual | 1.13 ** | -0.44 |
| | (0.41) | (0.48) |
| edgecov.hundreds_messages | 0.34 *** | 0.41 *** |
| | (0.09) | (0.09) |
| mix.leader.0.0 | -2.79 *** | -1.10 *** |
| | (0.16) | (0.21) |
| mix.leader.1.0 | -3.43 *** | -1.53 ** |
| | (0.48) | (0.56) |
| mix.leader.1.1 | -0.10 | -0.47 |
| | (0.43) | (0.54) |
| gwideg.fixed.0.693147180559945 | | -2.42 *** |
| | | (0.34) |
| gwodeg.fixed.2 | | -3.61 *** |
| | | (0.28) |
| gwesp.OTP.fixed.0.693147180559945 | | 0.64 *** |
| | | (0.10) |
| nodematch.female | | 0.19 |
| | | (0.14) |
| nodematch.department | | 2.11 *** |
| | | (0.18) |
| nodeicov.office | | -0.25 |
| | | (0.15) |
| nodeocov.office | | 0.03 |
| | | (0.19) |
| diff.t-h.tenure | | -0.14 *** |
| | | (0.02) |
| AIC | -708.62 | -1098.57 |
| BIC | -670.43 | -1015.83 |
| Log Likelihood | 360.31 | 562.28 |

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

2. **For each of the eight hypotheses, interpret the results from your models and state whether that hypothesis was supported. To determine whether a hypothesis is supported, look at whether there is a p-value < 0.05 and the directionality (positive/negative) of the effect. When you interpret the results, convert the model coefficients, which are given by R as conditional log-odds, into odds ratios.**

Answer:

All answers are based on the graph in question 1.

## Social Network Lab 2: Exponential Random Graph Models

*Student: Yijian Li   Net ID: YLS9426*

*Hypothesis 1: There will be indegree popularity effects – That is, a tendency for a small number of employees to be sought out for advice from many others (as opposed to advice seeking behaviors being spread evenly amongst all employees).*

For indegree popularity effects: we can see the parameter **gwidegree** in model 2.

Firstly, the p value is less than 0.001, so **it surely has impact on** our model/network, and with a **negative** value -2.42, we can know there is an indegree preferential attachment, **the hypothesis is true.**

**Exp(-2.42) = 0.08892162** => the tendency of an individual with less indegree value to be sought out for advice from many others are **0.08892162** times the tendency of an individual with many indegree value(indegree increase, tendency increase)

*Hypothesis 2: Individuals will be more likely to report go to advice from people in their own department, as opposed to other departments.*

For department tendency: we can see the parameter **nodematch.department** in model 2.

Firstly, the p value is less than 0.001, so **it surely has impact on** our model/network, and with a **positive** value 2.11, we can know there is a tendency of nodes to form ties with matching department values, **the hypothesis is true.**

**Exp(2.11) = 8.248241** => the tendency of an individual within the same department to be report for advice from many others are **8.248241** times the tendency of an individual in other departments.

*Hypothesis 3: There will be homophily based upon the sex of individuals, in terms of who employees go to for advice.*

For sex tendency: we can see the parameter **nodematch.female** in model 2.

As we can see, the p value is above 0.05, so we **cannot make our judgment** about whether the tendency of seeking for advice has a clear relationship with different sex categories in different individuals.

**The hypothesis cannot be determined.**

*Hypothesis 4: Employees who message someone more frequently on ESM will be more likely to report going to that person for advice.*

For ESM tendency: we can see the parameter **edgecov.hundreds_messages** in model 2.

Firstly, the p value is less than 0.001, so **it surely has impact on** our model/network, and with a **positive** value **0.41**, we can know there is covariance between out-degree of nodes and number of ESM of nodes, **the hypothesis is true.**

**Exp(0.41) = 1.506818** => the tendency of someone more frequency on ESM to be sought out for advice from many others are **1.506818** times the tendency of an individual with less frequency on ESM (More frequent on ESM, higher tendency)

**Social Network Lab 2: Exponential Random Graph Models**

*Student: Yijian Li   Net ID: YLS9426*

*Hypothesis 5: If an employee i goes to another employee j for advice, it will be more likely that j also goes to employee i for advice.*

For mutual effects: we can see the parameter **mutual** in model 2.

As we can see, the p value is above 0.05, so we **cannot make our judgment** about whether the tendency of seeking for advice has a clear relationship with number of reciprocal edges in the network.

**The hypothesis cannot be determined.**

*Hypothesis 6: Employees who work in the main office will be more likely to go to others for advice than employees from the secondary office.*

For main office outdegree tendency: we can see the parameter **nodeocov.office** in model 2.

As we can see, the p value is above 0.05, so we **cannot make our judgment** about whether the out-degree tendency of a person who are seeking for advice has a clear relationship with which office he/she belongs to.

**The hypothesis cannot be determined.**

*Hypothesis 7: Employees who work in the main office will be more likely to be sought after for advice than employees from the secondary office.*

For main office indegree tendency: we can see the parameter **nodeicov.office** in model 2.

As we can see, the p value is above 0.05, so we **cannot make our judgment** about whether the in-degree tendency of a person who are seeking for advice has a clear relationship with which office he/she belongs to.

**The hypothesis cannot be determined.**

*Hypothesis 8: Advice seeking relationships tend to be transitive - That is, if individual i goes to an individual k for advice, and k goes to an individual j for advice, then i is more likely to go to j for advice as well.*

For transitive effects: we can see the parameter **gwesp.OTP.fixed** in model 2.(model 2 is better than model 1)

Firstly, the p value is less than 0.001, so **it surely has impact on** our model/network, and with a **positive** value **0.65**, we can know there is **more likely to** have Directed Geometrically Weighted Outdegree attachment, **the hypothesis is true.**

# PART III: Model Diagnostics

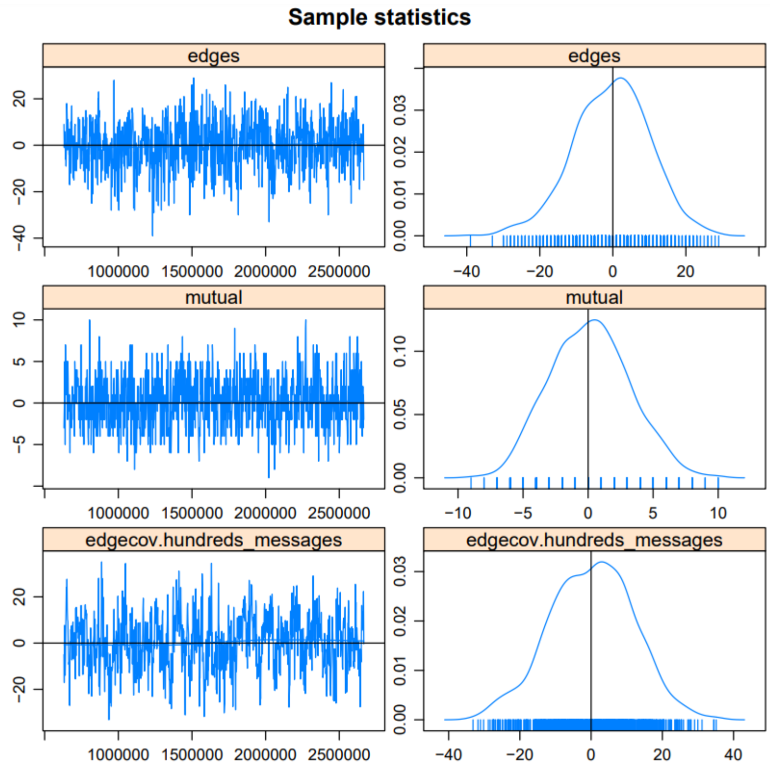
1. **Attach the model diagnostics for model 1 and 2 in your report (you should submit a single PDF file) and interpret the plots. Has the MCMC process converged to a desired state?**

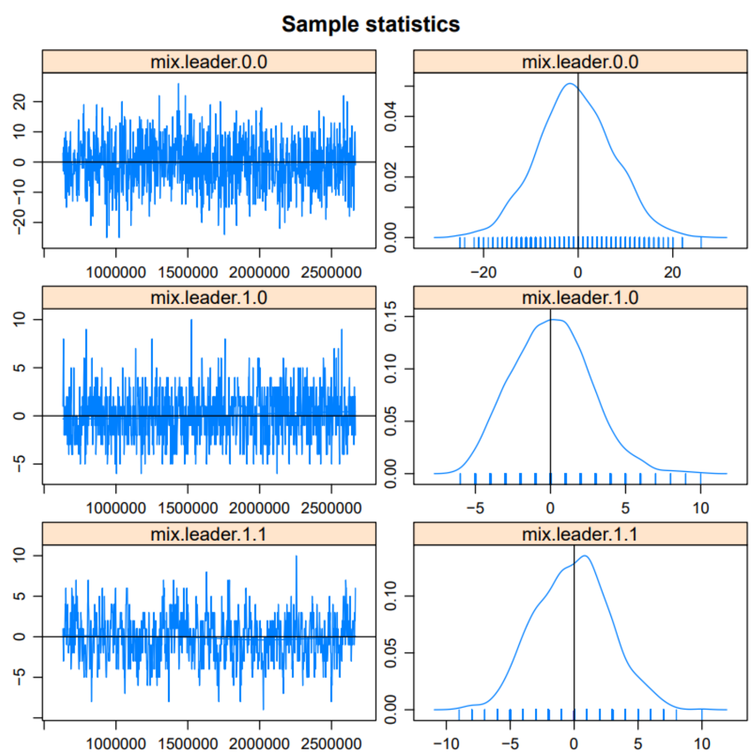
# Social Network Lab 2: Exponential Random Graph Models

*Student: Yijian Li   Net ID: YLS9426*

Answer:

## For model 1:

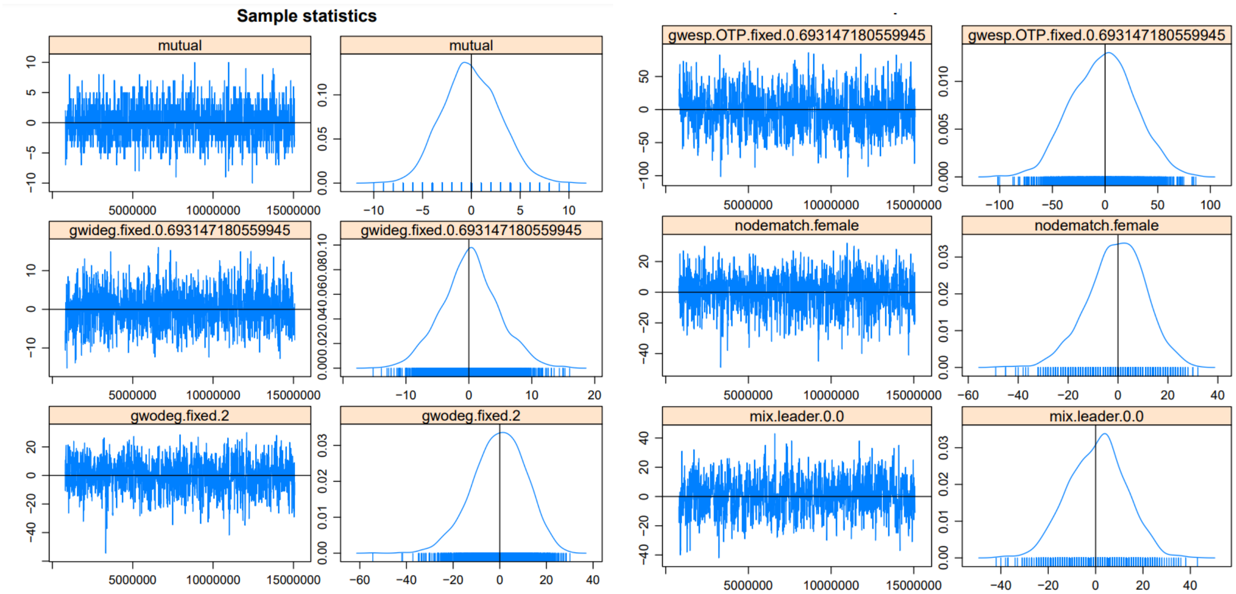


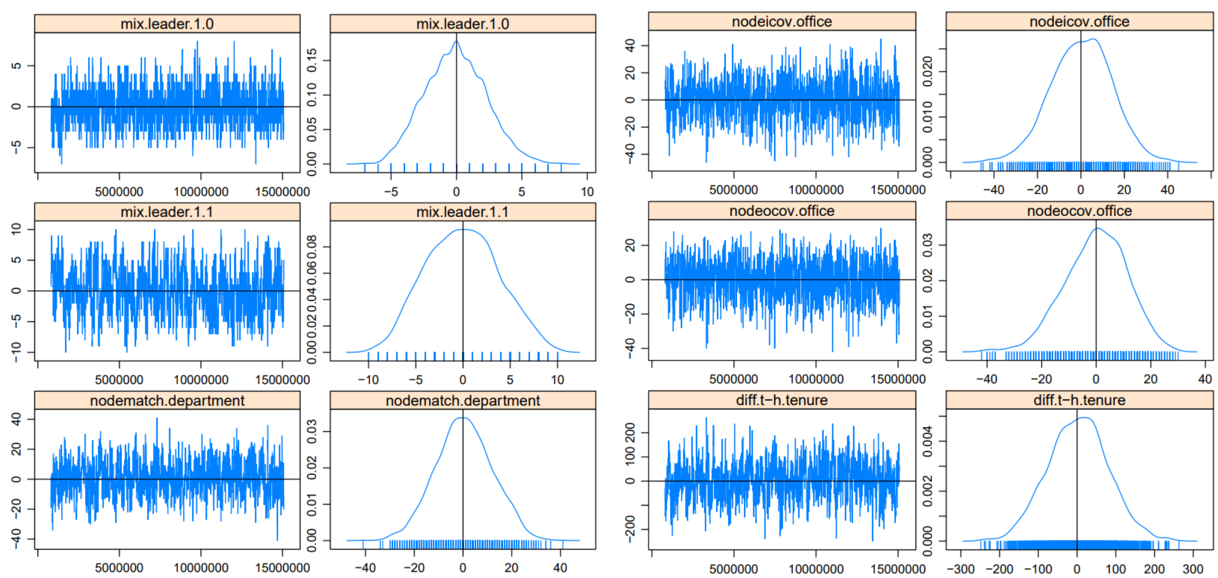



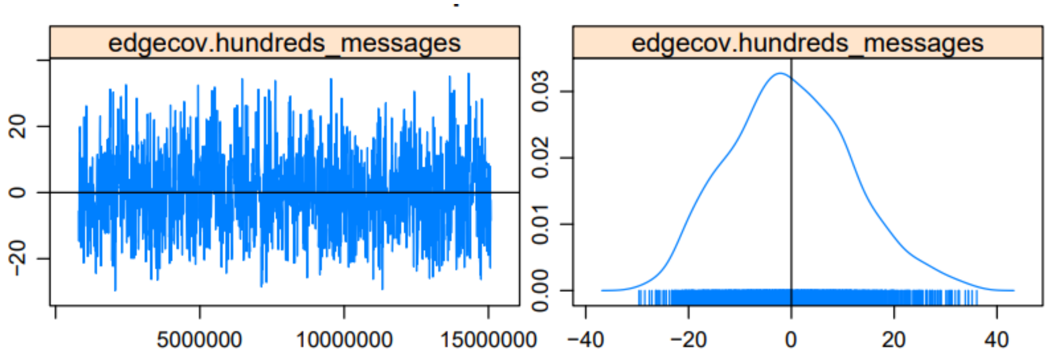
## Social Network Lab 2: Exponential Random Graph Models

*Student: Yijian Li   Net ID: YLS9426*

### For model 2:



## Social Network Lab 2: Exponential Random Graph Models

*Student: Yijian Li   Net ID: YLS9426*

All the graphs in model 1 and model 2 are good and **all converged** around a relatively stable value of the number of attributes.

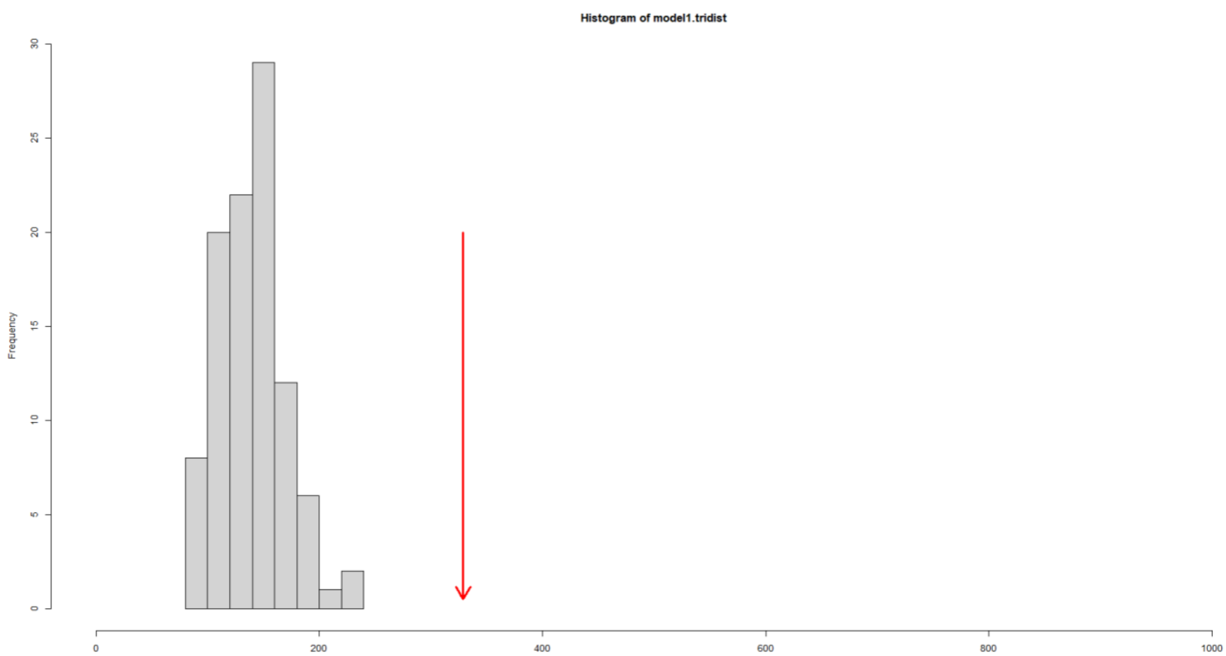
They are all in good shape and look like in normal distribution.

**Yes**, the MCMC process **converged** to a desired state as it should be.

2. **Perform Goodness of Fit test to check how well the estimated model captures certain statistical features of the observed network for both model 1 and 2.**
   a) **To do so, simulate many networks from the estimated model and extract 100 samples from the simulation process. Please note, this may take 2 minutes or more to compute.**
   b) **Extract the number of triangles from each of the 100 samples.**
   c) **Compare the distribution of triangles in the sampled networks with the observed network by generating a histogram of the triangles. Interpret your result -- is the estimated model a good one in terms of triangle measure?**

Answer:

**Model 1:**

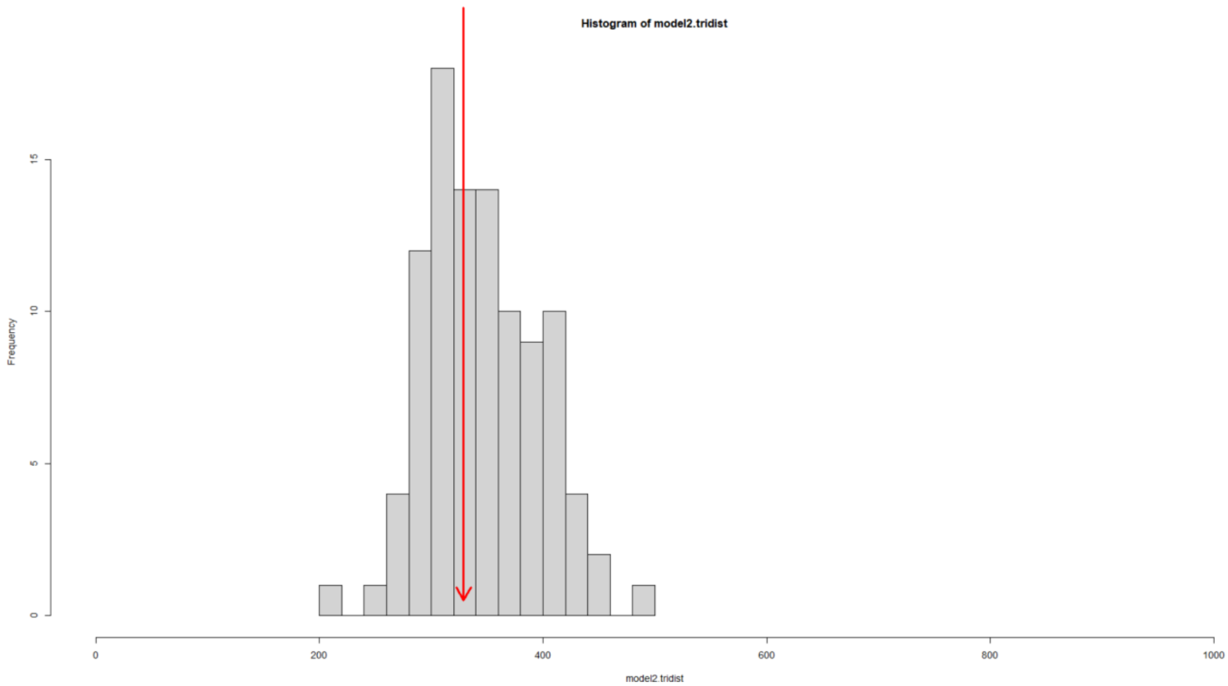



As we can see from the picture, **none** of our simulated networks have as many triangles as the actual advice network.

So, this model 1 might **not be a good reflection of reality** and not be controlling for the fact there are lots of triangles on the network.

**Social Network Lab 2: Exponential Random Graph Models**
*Student: Yijian Li   Net ID: YLS9426*

## Model 2:




Clearly, this graph looks much better.   We can see that the number of triangles in the real network **matches up** with the number of triangles in our simulations.

The model 2 is a **good one**.

3. **Repeat this goodness-of-fit evaluation process for a variety of other network statistics just for model 2 (for example, degree distribution, distribution of edgewise shared partners, and the distribution of geodesics). Simulate networks as we did above, compile statistics for these simulations as well as the observed network, and calculate p-values of all of the aforementioned values to evaluate the correspondence between the networks simulated by the model and the observed network. Report the p-values for the simulation and interpret them.**

Answer:

The model 2 is generally good, all those charts show that the model 2 **could fit** the real network. And thus, the **model 2 is a good model for us to use or implement.**

As we could see from the charts in the following:

**Social Network Lab 2: Exponential Random Graph Models**

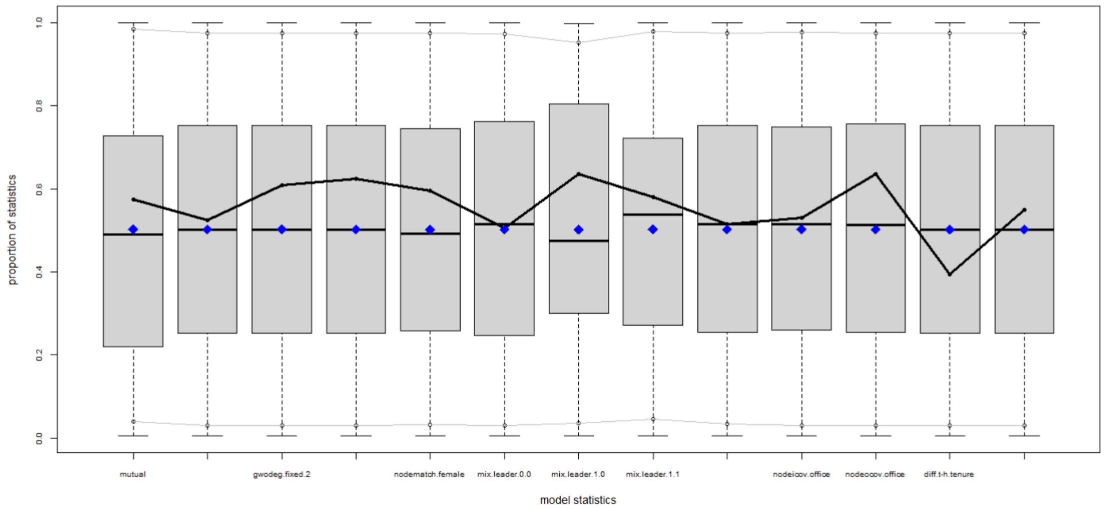
*Student: Yijian Li Net ID: YLS9426*

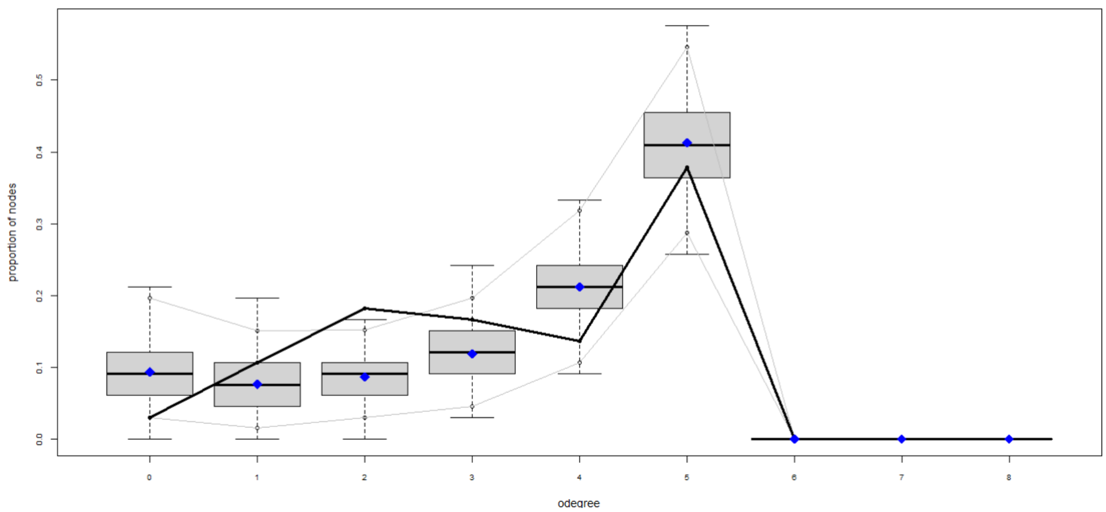
Basically, the model has minimum p-value as 0.84, and maximum as 1.00. They are all greatly more than 0.05, so we know they are all insignificant to the network, which also indicates that this is a good fit for us. The model simulated by model2 is corelated to the observed network and this model2 is the appropriate model we could use in the further analysis.

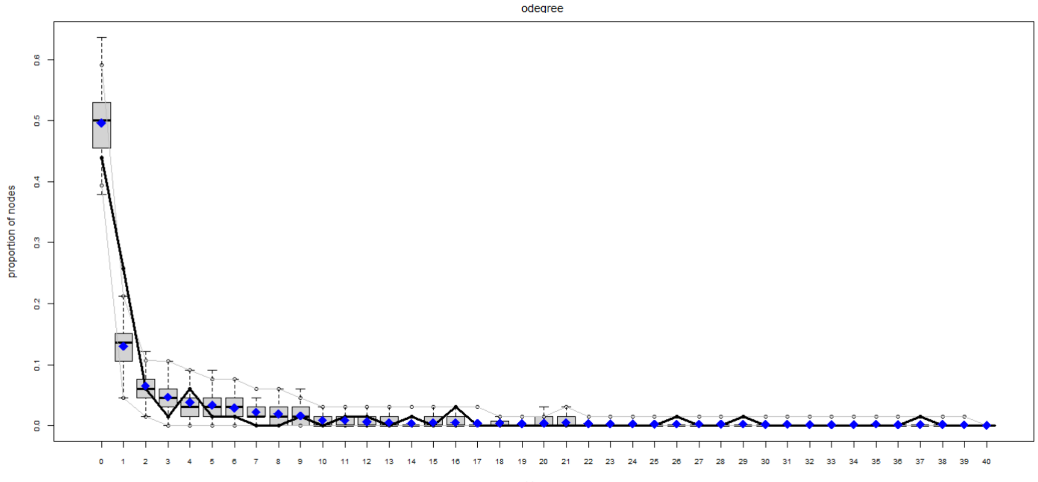
**Model 2**

**Social Network Lab 2: Exponential Random Graph Models**

*Student: Yijian Li   Net ID: YLS9426*

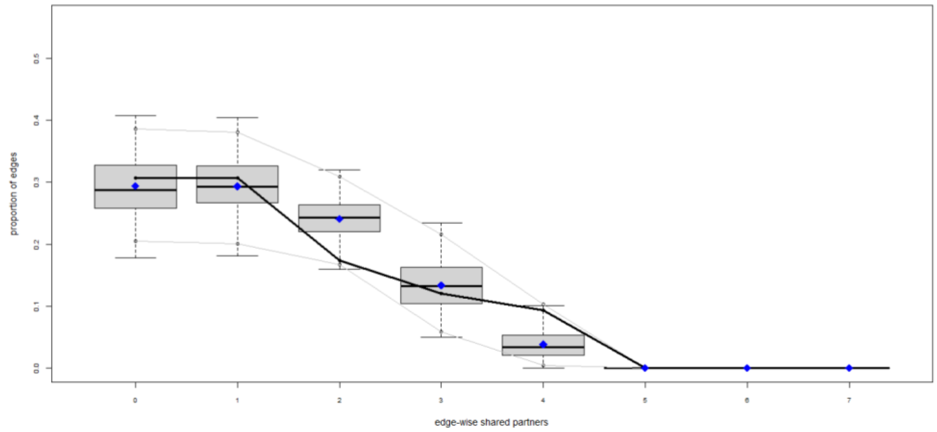


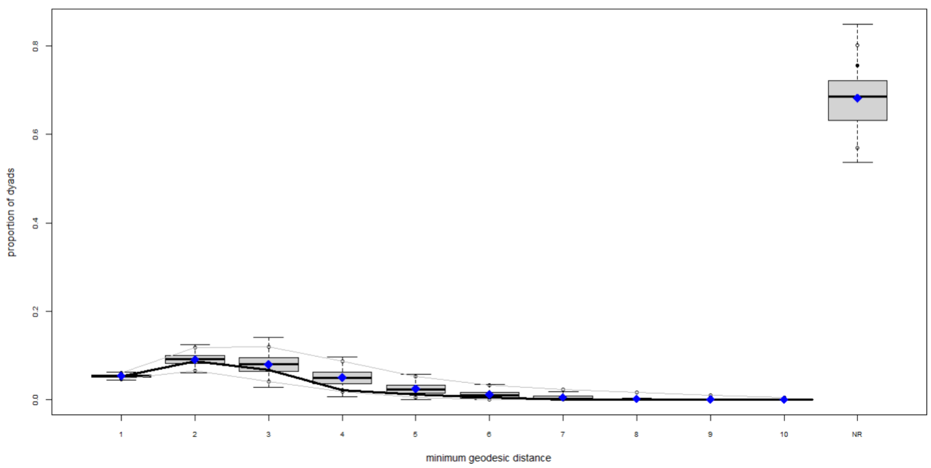



# Social Network Lab 2: Exponential Random Graph Models

*Student: Yijian Li   Net ID: YLS9426*




Goodness-of-fit diagnostics




Goodness-of-fit for model statistics

| | obs | min | mean | max | MC p-value |
|---|---|---|---|---|---|
| mutual | 12.0000 | 5.0000 | 11.75500 | 21.00000 | 1.00 |
| gwideg.fixed.0.693147180559945 | 54.1507 | 40.2011 | 53.82764 | 64.98289 | 0.98 |
| gwodeg.fixed.2 | 182.8840 | 145.8688 | 183.87301 | 208.03749 | 0.86 |
| gwesp.OTP.fixed.0.693147180559945 | 214.1250 | 136.0000 | 216.56062 | 280.25000 | 0.85 |
| nodematch.female | 162.0000 | 122.0000 | 162.78500 | 189.00000 | 0.93 |
| mix.leader.0.0 | 94.0000 | 60.0000 | 91.54000 | 128.00000 | 0.90 |
| mix.leader.1.0 | 8.0000 | 2.0000 | 8.04500 | 14.00000 | 1.00 |
| mix.leader.1.1 | 12.0000 | 3.0000 | 12.91000 | 22.00000 | 0.89 |
| nodematch.department | 112.0000 | 81.0000 | 109.64000 | 145.00000 | 0.84 |
| nodeicov.office | 182.0000 | 134.0000 | 180.52500 | 230.00000 | 0.96 |
| nodeocov.office | 173.0000 | 129.0000 | 173.54000 | 200.00000 | 0.98 |
| diff.t-h.tenure | -737.2575 | -909.1479 | -747.50245 | -510.88493 | 0.86 |
| edgecov.hundreds_messages | 56.6900 | 30.3900 | 57.58930 | 86.59000 | 0.95 |