1. First, we want to know if network familiarity influences an investor choosing to invest in a particular startup executive. ERGM models allow us to estimate a regression in which the outcome variable is the network itself. We would like to estimate the relationship, “chooses to invest in,” which can be represented as a directed network from funding PersonIds to funded PersonIds. Considerinvestmentsto beallfinancingdealsoutsideofBuyouts, Mergers/Acquisitions, and IPOs. To make sure that the nodes in the network are likely to be drawn from a more established community, limit the analysis to only deals in which all participating investors are from US cities with at least 1,000 investor firms. For similar reasons, and to limit computational strain, also only consider deals from the year 2000 onward. Run an ERGM model, using 20 iterations, that predicts the likelihood of a “chooses to invest in” relationship as a function of the edges in a model, as well as the presence of triangles, which represent the influence of “friends of friends” on the likelihood of funding. What do the results suggest about potential network familiarity effects in investment decisions?

Answer:

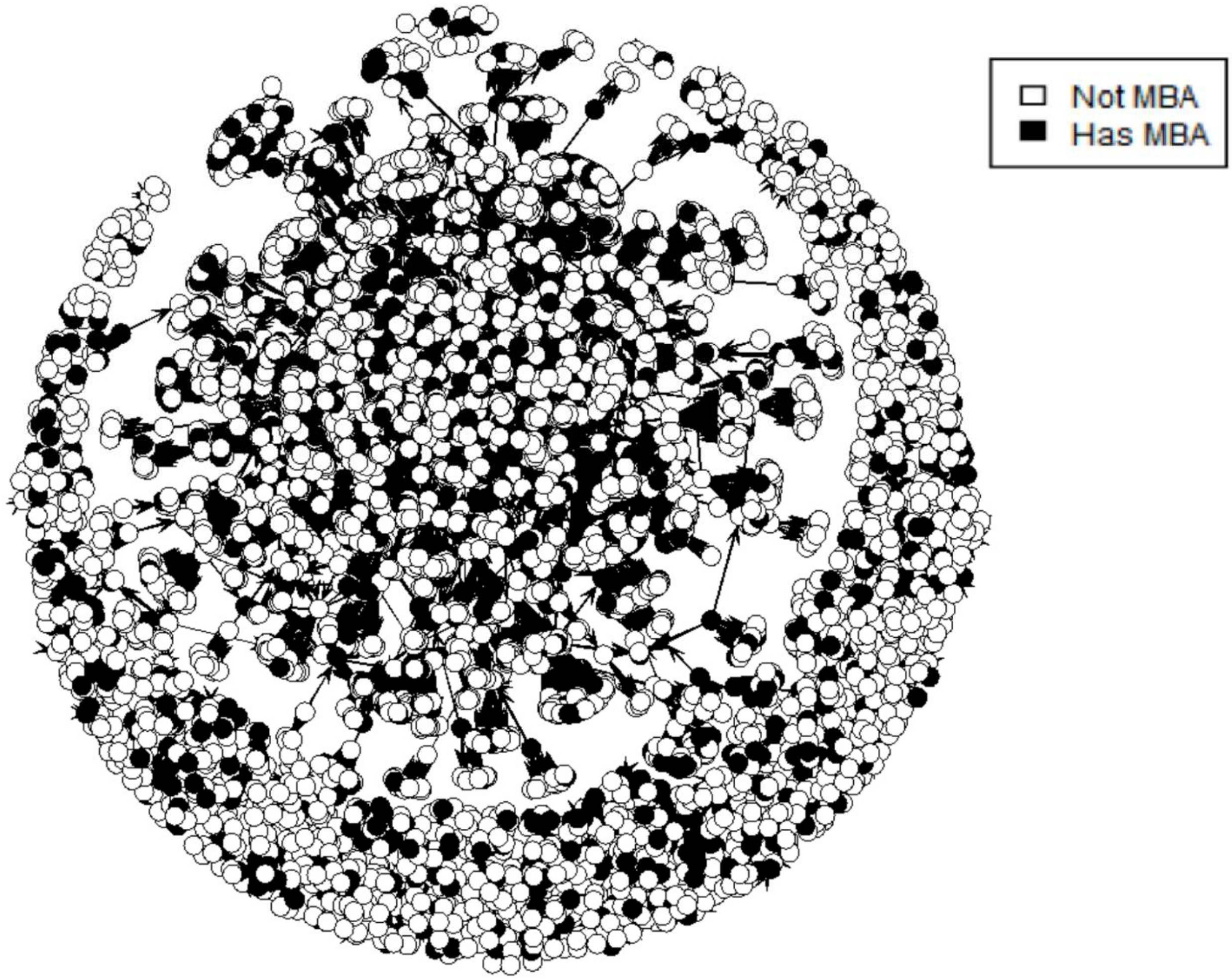
As suggested, I am selecting *"Los Angeles"* which has more than 1000 investors. The deals after 2000 are only considered for this assignment. I have also cleaned the deal types as suggested in the question.

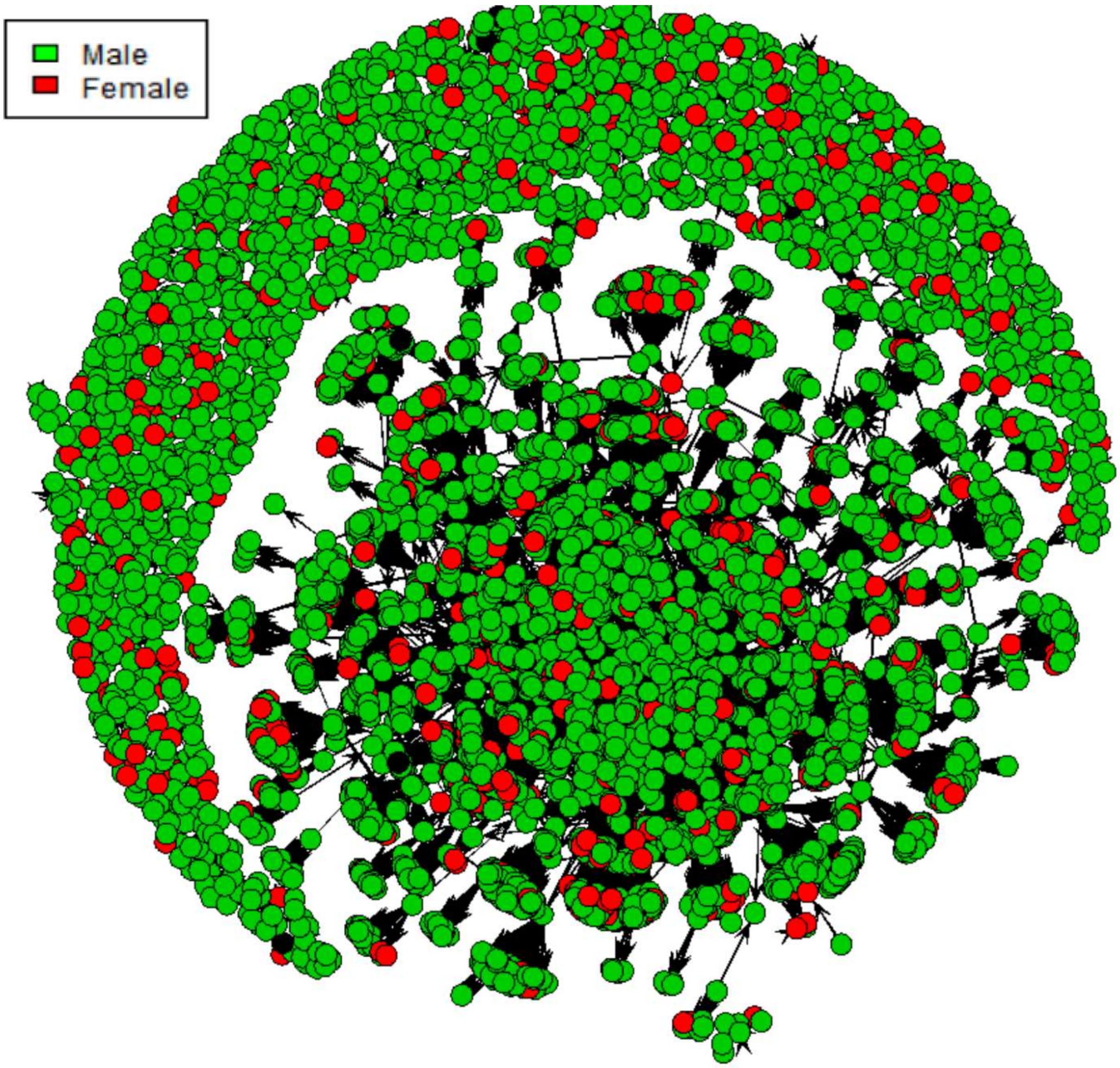
The results suggest that the *“friends of friends”* influences the edges presence. This means the triangle is significant estimator in the created network of “chooses to invest in” investing person to start up executive. If the person chooses to invest in an executive and same investor is investing in some other executive, our network show triangle relation similar to *“friends of friends”* network.

```
==========================
Summary of model fit
==========================

Formula:   n ~ edges + triangle

Iterations:  20 out of 20

Monte Carlo MLE Results:
         Estimate Std. Error MCMC % z value Pr(>|z|)
edges    -8.54938    0.01570      1  -544.6   <1e-04 ***
triangle  5.02370    0.04395     11   114.3   <1e-04 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

     Null Deviance: 41063156  on 29620806  degrees of freedom
 Residual Deviance:   131137  on 29620804  degrees of freedom

AIC: 131141    BIC: 131171    (Smaller is better.)
```

2. Next, we want to include information on the investors and executives to determine whether there is homophily in choosing to make investment decisions. Run an ERGM model, using 20 iterations, that predicts the likelihood of a “chooses to invest in” relationship as a function of the edges in a model, as well as the presence of triangles. This time, also include the effect of having the same gender and of the investor and the executive both having an MBA. What do the results suggest about homophily in investment decisions?

Answer:


Not MBA
Has MBA


Male
Female

```
==========================
Summary of model fit
==========================

Formula:   invest_net ~ edges + triangle + nodematch("Gender", diff = T) +
    nodematch("has_MBA", diff = T)

Iterations:  20 out of 20

Monte Carlo MLE Results:
                        Estimate Std. Error MCMC %  z value Pr(>|z|)
edges                   -8.83643    0.03536      1 -249.898  < 1e-04 ***
triangle                 5.16055    0.02471     13  208.867  < 1e-04 ***
nodematch.Gender.Female -1.12863    0.23302      1   -4.844  < 1e-04 ***
nodematch.Gender.Male    0.38366    0.03611      1   10.625  < 1e-04 ***
nodematch.has_MBA.FALSE -0.11393    0.03138      1   -3.630 0.000283 ***
nodematch.has_MBA.TRUE   0.21358    0.06729      1    3.174 0.001505 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

     Null Deviance: 41063156  on 29620806  degrees of freedom
 Residual Deviance:   130706  on 29620800  degrees of freedom

AIC: 130718    BIC: 130809    (Smaller is better.)
```

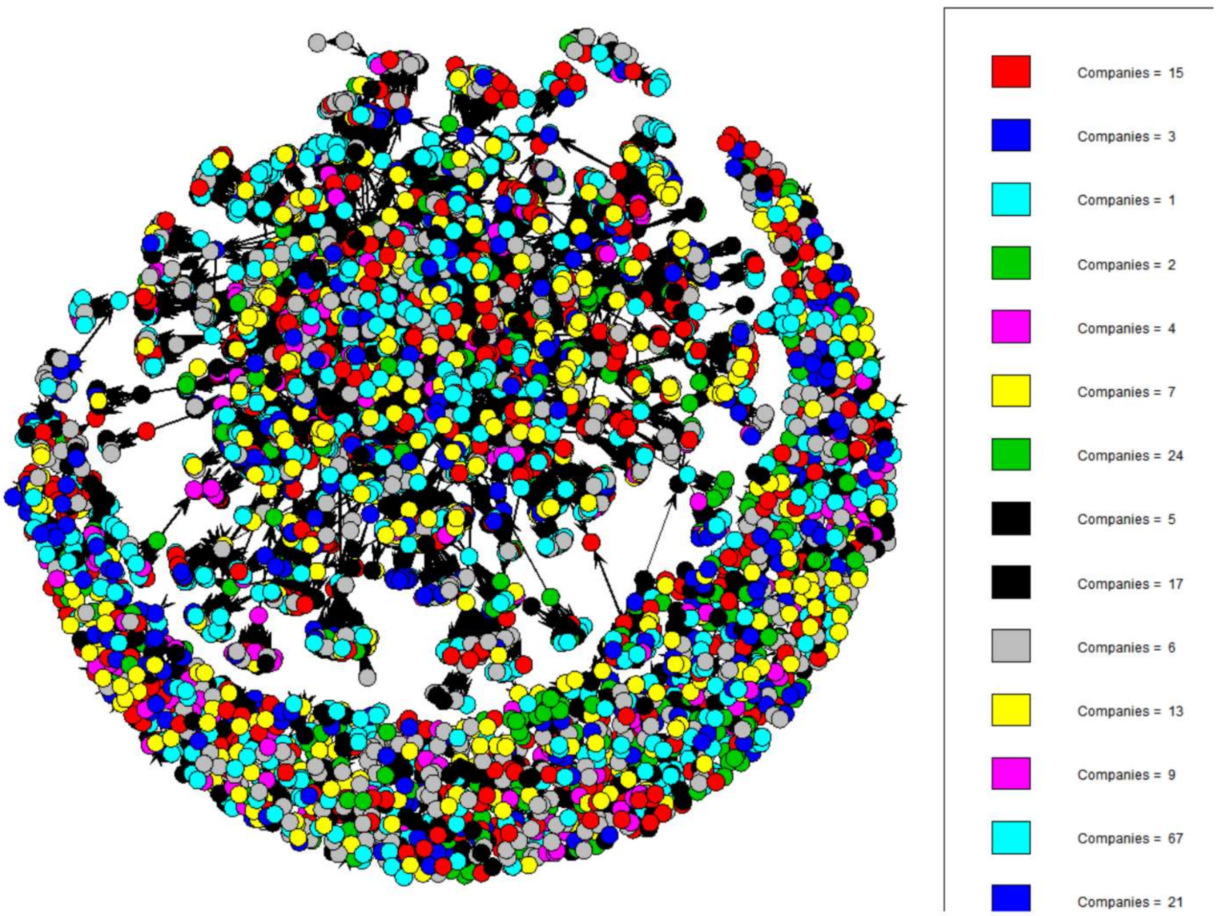
The above results show that there is homophily in case of gender and MBA. That MBA investors have more probability of connections with MBA's. From the available network it can be seen that Males form an homophily in the network. MBA's invest in another MBA's. Males invest in male executive in terms of more likelihood.

3. Last,wewanttoknowhowtheseresultsmightbeaffectedifweincludeinformationaboutthe quality of the startup executives. If investors are choosing based on merit, when we include quality metrics for the entrepreneurs these should dominate over any other predictors. Run an ERGM model, using 20 iterations, that predicts the likelihood of a "chooses to invest in" relationship as a function of the edges in a model, the presence of triangles, the effect of having the same gender, and of the investor and the executive both having an MBA. This time, also include the total number of different companies the startup investor has worked for,aswellasthetotalnumberofsuccessfuldealstheexecutivehasbeenapartof. Successful deals can include Buyouts, Mergers/Acquisitions, and IPOs. What do these new results suggest about what drives investors' decisions to invest in an entrepreneur or not? Is this problematic for these industries?

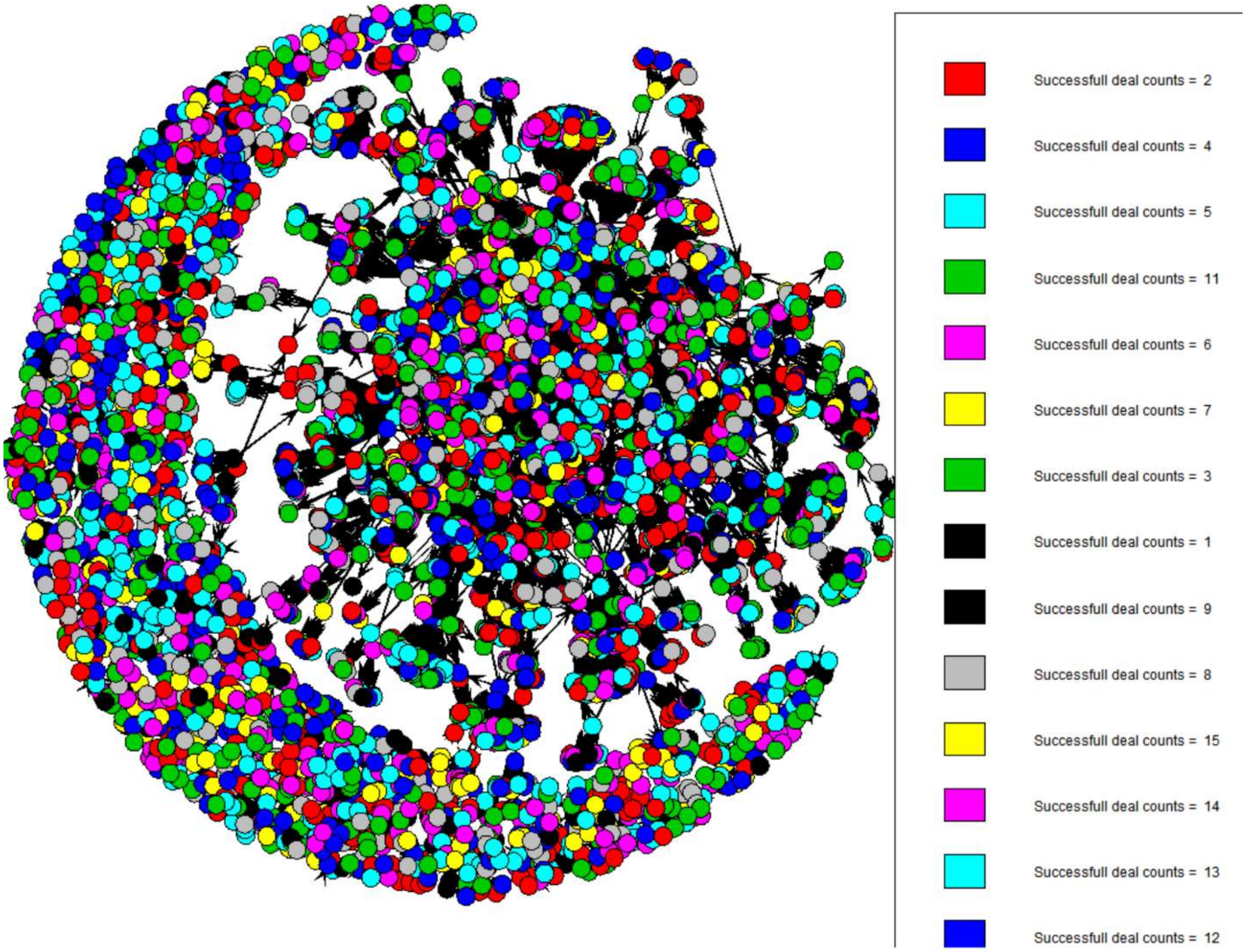
Answer:

For this question we will consider all the deals including the ones filtered for pevious questions.

Network for quality as number of different companies the investor has invested in.



Network for quality as number of successful deals.




```
==========================
Summary of model fit
==========================

Formula:   invest_net ~ edges + triangle + nodematch("Gender", diff = T) + 
    nodematch("has_MBA", diff = T) + nodecov("successful_deals_cnt") + 
    nodecov("diff_Company_Cnt")

Iterations:  20 out of 20 

Monte Carlo MLE Results:
                               Estimate Std. Error MCMC %  z value Pr(>|z|)    
edges                        -8.5727229  0.0437829      1 -195.801   <1e-04 ***
triangle                      1.7010927  0.0048334      4  351.945   <1e-04 ***
nodematch.Gender.Female      -1.2247951  0.1968698      2   -6.221   <1e-04 ***
nodematch.Gender.Male         0.4619019  0.0374649      1   12.329   <1e-04 ***
nodematch.has_MBA.FALSE      -0.2067429  0.0323362      1   -6.394   <1e-04 ***
nodematch.has_MBA.TRUE        0.3772610  0.0626563      1    6.021   <1e-04 ***
nodecov.successful_deals_cnt -0.0098581  0.0008910      1  -11.065   <1e-04 ***
nodecov.diff_Company_Cnt      0.0060492  0.0006336      1    9.548   <1e-04 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

     Null Deviance: 41063156  on 29620806  degrees of freedom
 Residual Deviance:   128822  on 29620798  degrees of freedom
 
AIC: 128838    BIC: 128960    (Smaller is better.)
```

The above results suggest that the quality/merit of the person is also an import factor in edge formation. The standard error is less if we get different company counts and successful deals counts. If the person has more number of successful deals, he is more likely to receive more number of investments. The quality variable dominates other factors.

Extra Credit: Dynamic Analysis of familiarity versus investor quality—5 points

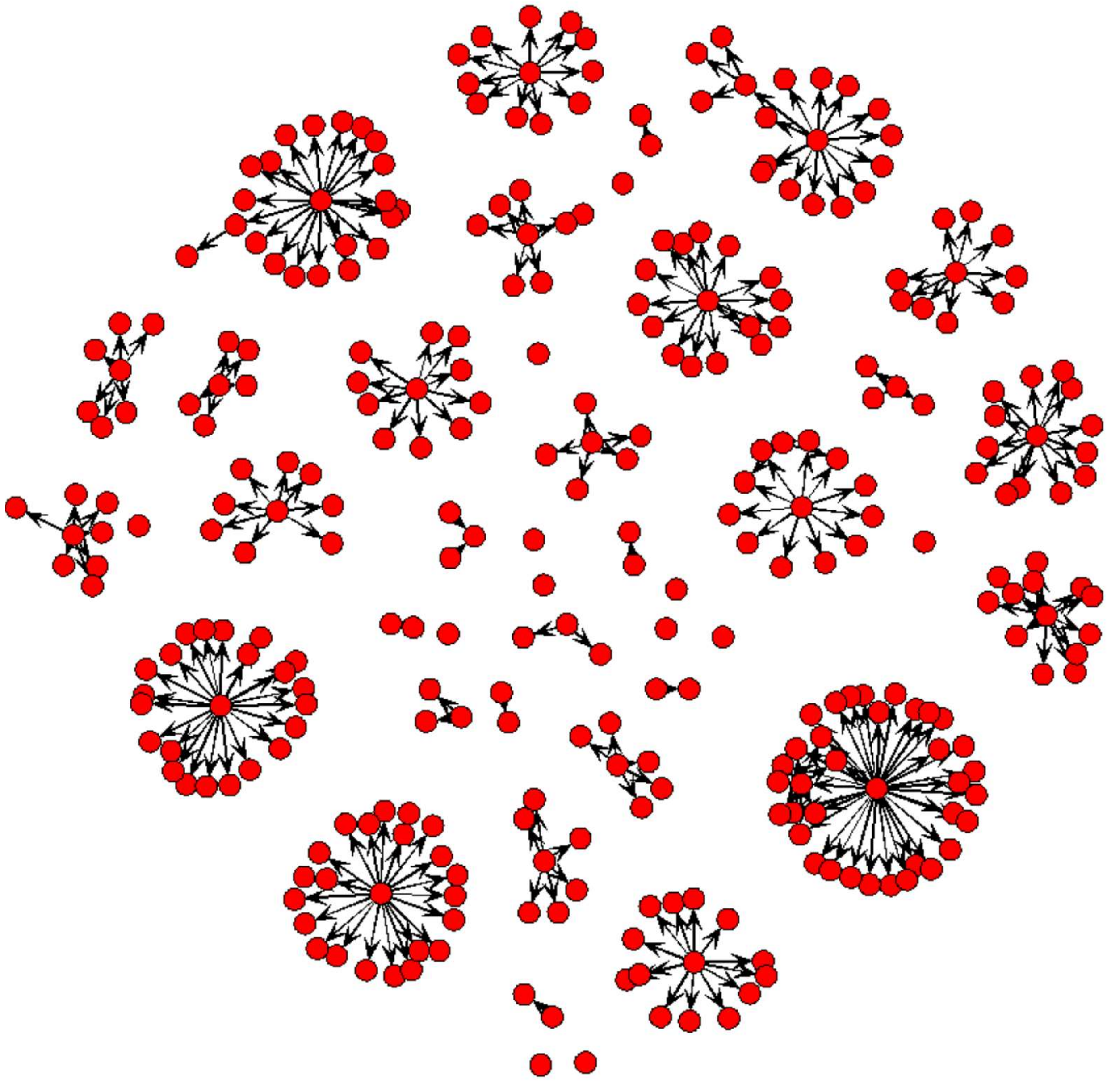
Thedataonindividualinvestorsandstartupexecutivesgrowsovertime. Wecantakeadvantageof thisfeaturetorunamorepowerfulsetofmodelsfromtheSienafamily,whichusesthelongitudinal nature of the data to predict how the network evolves over time. Benefits of the Siena model are that it can predict both ties as well as behaviors, and can take into account nodes joining and leaving the network over time. The model also accounts for ties not reappearing in future periods, rather than assuming that they last forever, so it gives a richer sense of the dynamic decisionmaking of individuals in the network. We did not have time to cover Siena in detail in class, but it would be an applicable model for this data. The example script “rsiena teenage drug and alcohol usage.R” walks through an example of using the Siena modeling technique on a dataset that predicts smoking and alcohol consumption at a high school based on the network connections and behavioral similarity of the school’s students. The data for this analysis are included as well. You can adapt the approach to an analysis of the investor to startup executive network as well. To set up time windows for analysis that will not be too computationally intensive, you can use waves, rather than years, grouping the firms into three waves from 2000-2018. Be creative in your Siena modeling, and see what interesting effects you can find. The algorithm can take some time to run, especially if it needs to be re-run in order to converge, so you may want to run it overnight or during idle time. For Mac users, remember that you can open multiple instances of any application by opening the terminal and typing in “open -n” and the name of the application; for example “open -n “R””.

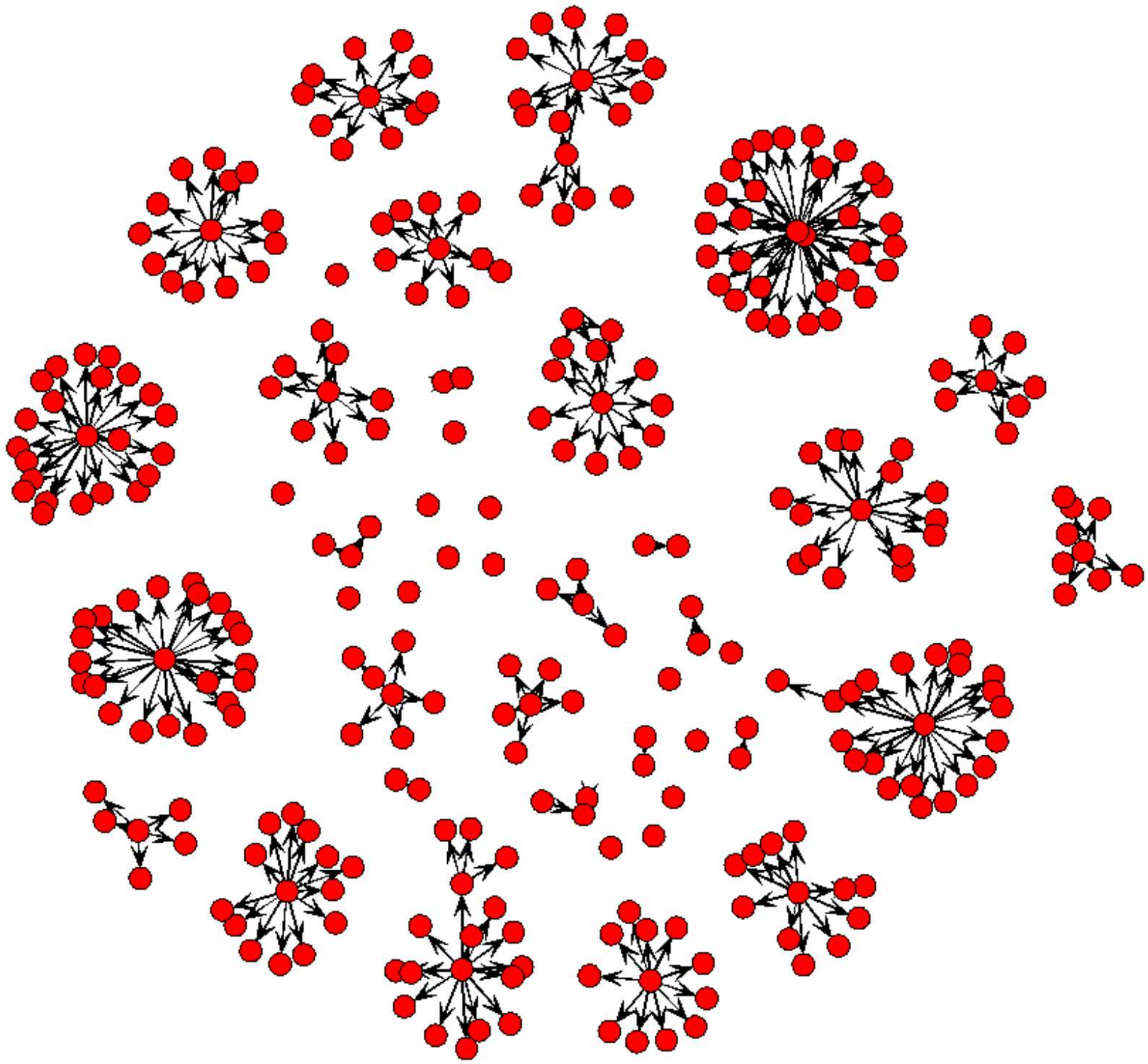
I have considered 3 cities for my network “Chicago, San Francisco, and New Jersey” This will limit the network. I created common nodes for all three waves, which comes out to be close to 1000.

I have considered 3 waves. Dates for creating waves are cut as follows
Wave1 < '2013-01-01',
'2013-01-01' <= Wave 2 < '2015-01-01',
Wave3 >= '2015-01-01'

Behavior I selected for here is “deal size”

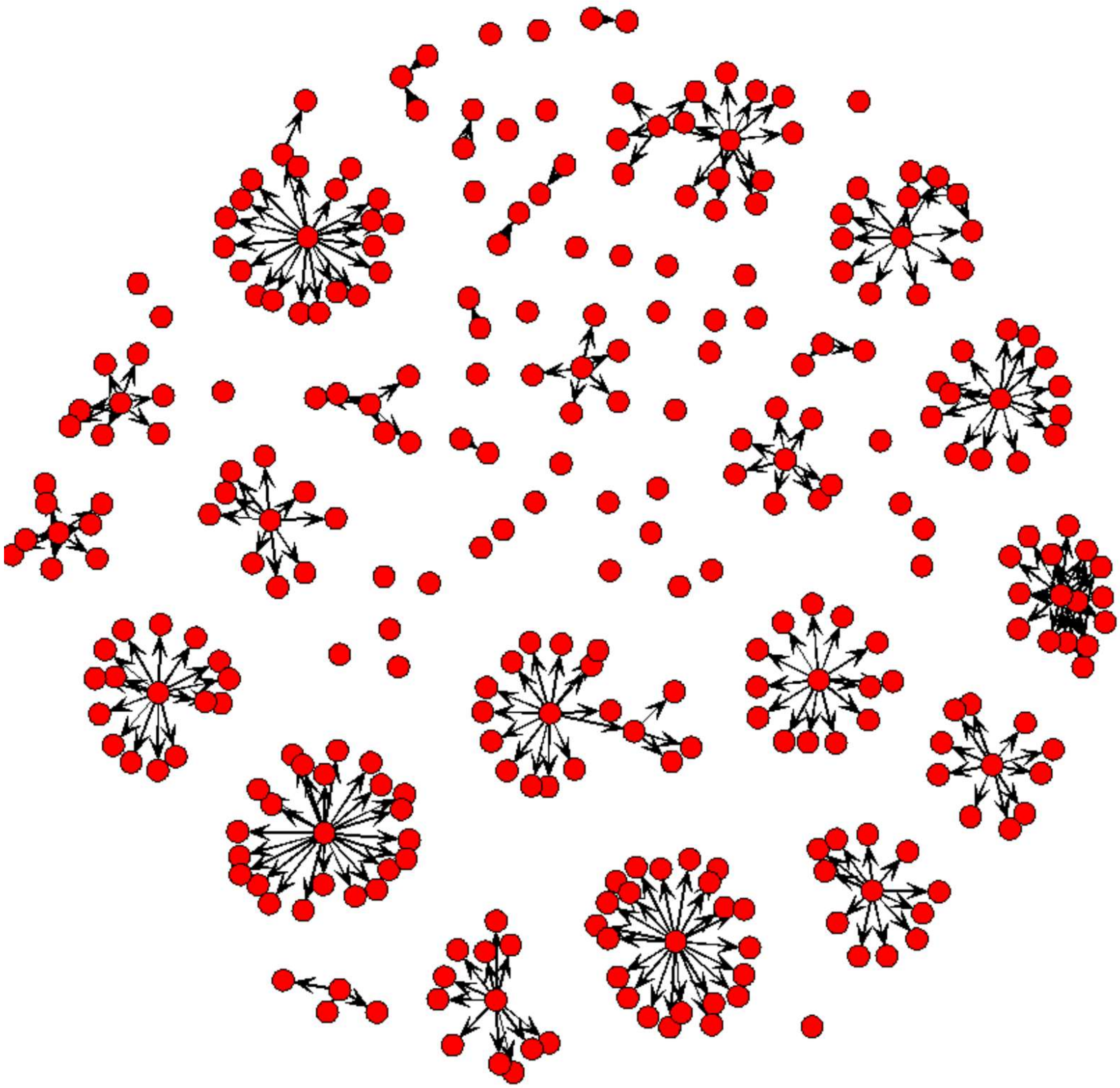
WAVE 1





Wave 2

WAVE 3



WAVE 3

The test for the outdegree effect comes close to significance, the other two are not significant.

The investment and deal sizes are two interactions.

The number of epochs were exceeding 1 million, Siena gave error. I changed the waves to reduce the size of network. The same things can be run on bug network if the hardware architecture is available. I reduced the network to only 11 nodes to get the results.

Siena Results:

```
                              ************************
                                       s50_3_CoEvinit_investments.out
                              ************************

Filename is s50_3_CoEvinit_investments.out.

This file contains primary output for SIENA project <<s50_3_CoEvinit_investments>>

Date and time: 05/12/2018 7:41:39 PM

RSiena version 1.2-12 (12 05 2018)


@1
Data input.
===========

3 observations,
11 actors,
1 dependent network variables,
0 dependent bipartite variables,
1 dependent behavior variables,
0 constant actor covariates,
0 exogenous changing actor covariates,
0 constant dyadic covariates,
0 exogenous changing dyadic covariates,
0 no files with times of composition change.


@2
Reading network variables.
--------------------------

Name of network variable: investments_dep.
oneMode network.
For observation moment 1, degree distributions are as follows:
Nodes
 1  2  3  4  5  6  7  8  9 10 11
out-degrees
 0  6  1  0  0  0  0  0  0  0  0
in-degrees
 0  0  0  0  1  1  1  1  1  1  1

No missing data for observation 1.

For observation moment 2, degree distributions are as follows:
Nodes
 1  2  3  4  5  6  7  8  9 10 11
out-degrees
 1  6  0  0  0  0  0  0  0  0  0
in-degrees
 0  0  0  1  1  1  1  1  1  1  0
```

```
No missing data for observation 2.

For observation moment 3, degree distributions are as follows:
Nodes
 1  2  3  4  5  6  7  8  9 10 11
out-degrees
 1  0  1  0  0  0  0  0  0  0  0
in-degrees
 0  0  0  1  0  0  0  0  0  0  1

No missing data for observation 3.





@2
Reading dependent actor variables.
----------------------------------

1st dependent actor variable named deal_size_dep.
Maximum and minimum rounded values are 0 and 20.
Non-integer values noted in this behavior variable: they will be truncated.


A total of 1 dependent actor variable.

Number of missing cases per observation:
 observation         1         2         3      overall
deal_size_dep         0         0         0         0     ( 0.0 %)

Means per observation:
 observation         1         2         3      overall
deal_size_dep      1.705     1.691     3.818     2.405



Behavior variable deal_size_dep:
All behavior changes are upward for the following periods:
2 => 3;
This will be respected in the simulations.

The mean structural dissimilarity value subtracted in the
balance calculations is         0.1273.

For the similarity variable calculated from each actor covariate,
the mean is subtracted.
These means are:
Similarity deal_size_dep            :       0.8379
```

```
@1
Initial data description.
==========================


@2
Change in networks:
--------------------

For the following statistics, missing values (if any) are not counted.

Network density indicators:
observation time                1      2      3
density                     0.064  0.064  0.018
average degree              0.636  0.636  0.182
number of ties                  7      7      2
missing fraction            0.000  0.000  0.000

The average degree is 0.485


Tie changes between subsequent observations:
 periods         0 =>  0    0 =>  1    1 =>  0    1 =>  1   Distance Jaccard   Missing
  1 ==>   2        102          1          1          6          2     0.750         0 (0%)
  2 ==>   3        102          1          6          1          7     0.125         0 (0%)

Directed dyad Counts:
 observation    total    mutual    asymm.     null
     1.            110        0        14        96
     2.            110        0        14        96
     3.            110        0         4       106

Standard values for initial parameter values
--------------------------------------------------

constant investments_dep rate (period 1)      0.4162
constant investments_dep rate (period 2)      1.4072
outdegree (density)                               -1.5624



@2
Dependent actor variables:
---------------------------

deal_size_dep
```

```
@3
Marginal distribution

                    Observations
values                     1      2      3
                    ------------------
  0                        9      8      8
  1                        0      1      0
  2                        0      1      0
  3                        1      0      0
  4                        0      0      1
  5                        0      0      0
  6                        0      0      0
  7                        0      0      0
  8                        0      0      0
  9                        0      0      0
 10                        0      0      0
 11                        0      0      0
 12                        0      0      0
 13                        0      0      0
 14                        0      0      0
 15                        1      1      0
 16                        0      0      0
 17                        0      0      0
 18                        0      0      1
 19                        0      0      0
 20                        0      0      1
No missings



@3
Changes

 periods    actors:  down   up   constant  missing  ;   steps:   down    up  total
  1  =>  2             1     1       9         0                 2.65   2.5  5.15
  2  =>  3             0     3       8         0                    0  23.4  23.4

For this variable, the standard initial behavioral tendency parameter is   0.0664

Initialisation of project <<s50_3_CoEvinit_investments>> executed succesfully.
```