DS745 Network Analysis

Telling a Network Story: Global Online Course Social Interactions

Analysis using Exponential Random Graph Modelling (ERGM) and Network Visualization

Rick W. Lentz, MS, MS

29 February, 2018 - BEST VIEWED HERE

Dataset: Explanation

After reviewing the publically available datasets, I found the CMOOC dataset interesting since it captured social interactions between 3 teachers and 764 students of an distributed massive open online course offered in 2011. A link was added whenever party A referenced part B in their Tweet.

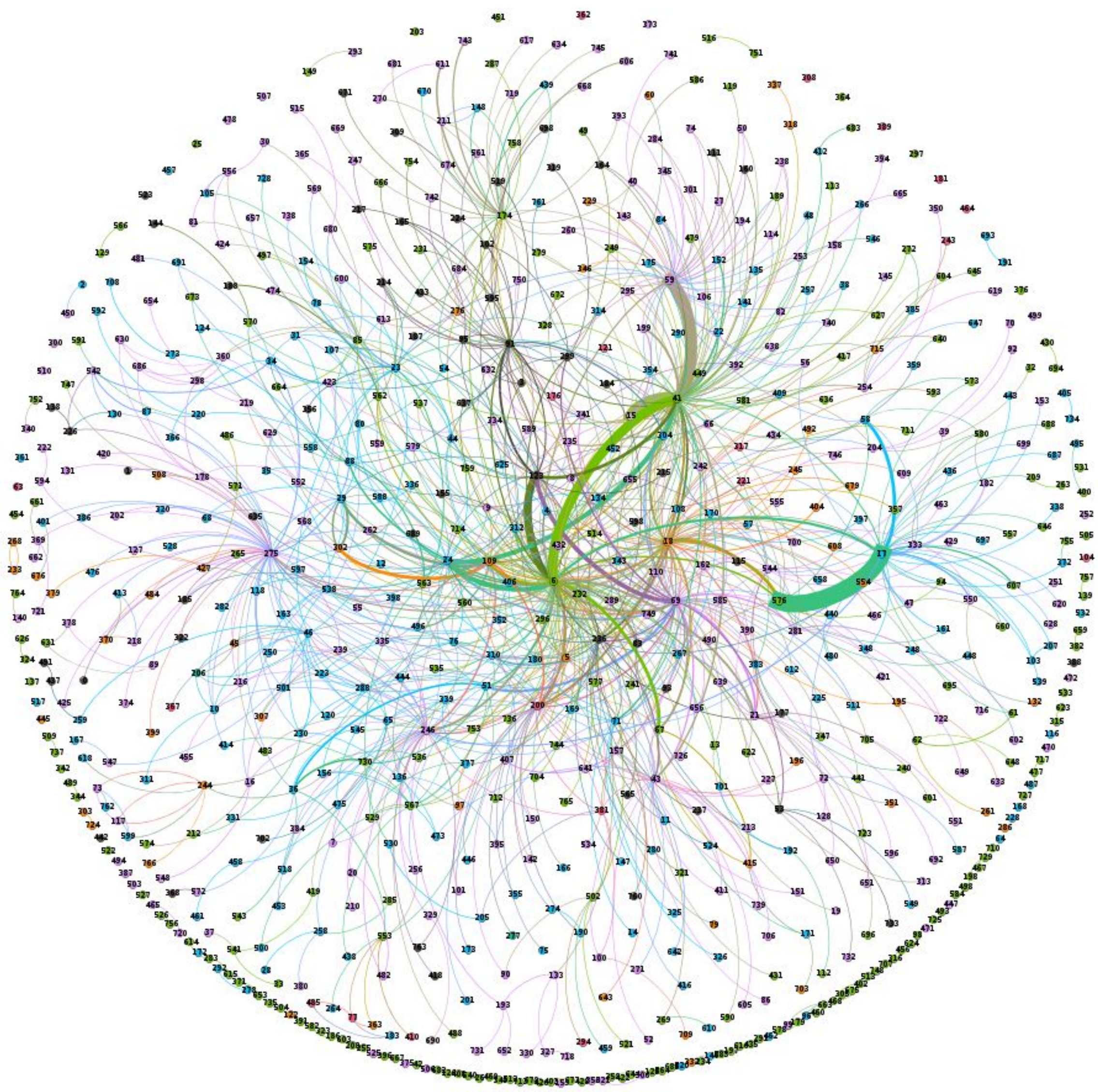
To perform the Exponential random graph model (ERGM) analysis, common elements of the object definition are extracted and added as node level attributes. Here is a table detailing the network's vertex attributes.

Attribute Name	Example	Description
id	123	Subject ID, range 1-767
domain	Undergraduate, Community	Social cluster, group
gender	M, F, unk/other, Org	Reported gender or organization or unk
role	Instr, Student	Role in network
continent	Intl, Asia, Europe	Reported continent
socio_tech	Social, Technical, Unk	Category Social or Techincal, or Unknown
work_type	Mixed, Practice, Research	Employment role or Unknown

Visualization and Network Summary

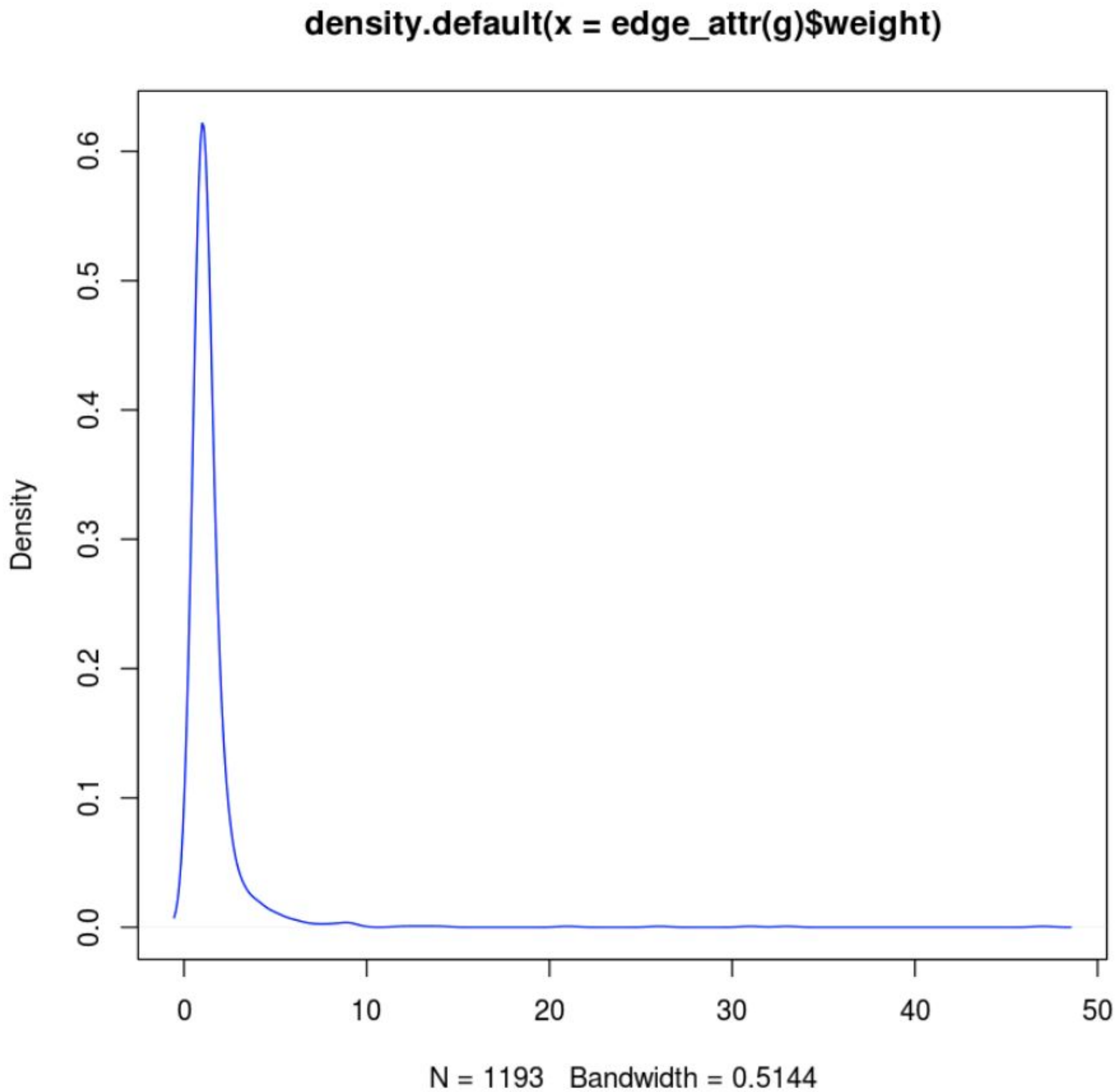
We start by looking at the overall color-coded structure of the entire network. Then analysis is shown for the network's structural properties. These will help build up ideas about the features we may want to include in models created in later sections of this paper.

The network consists of 1193 interactions (edges) and has a small density value of 0.002.

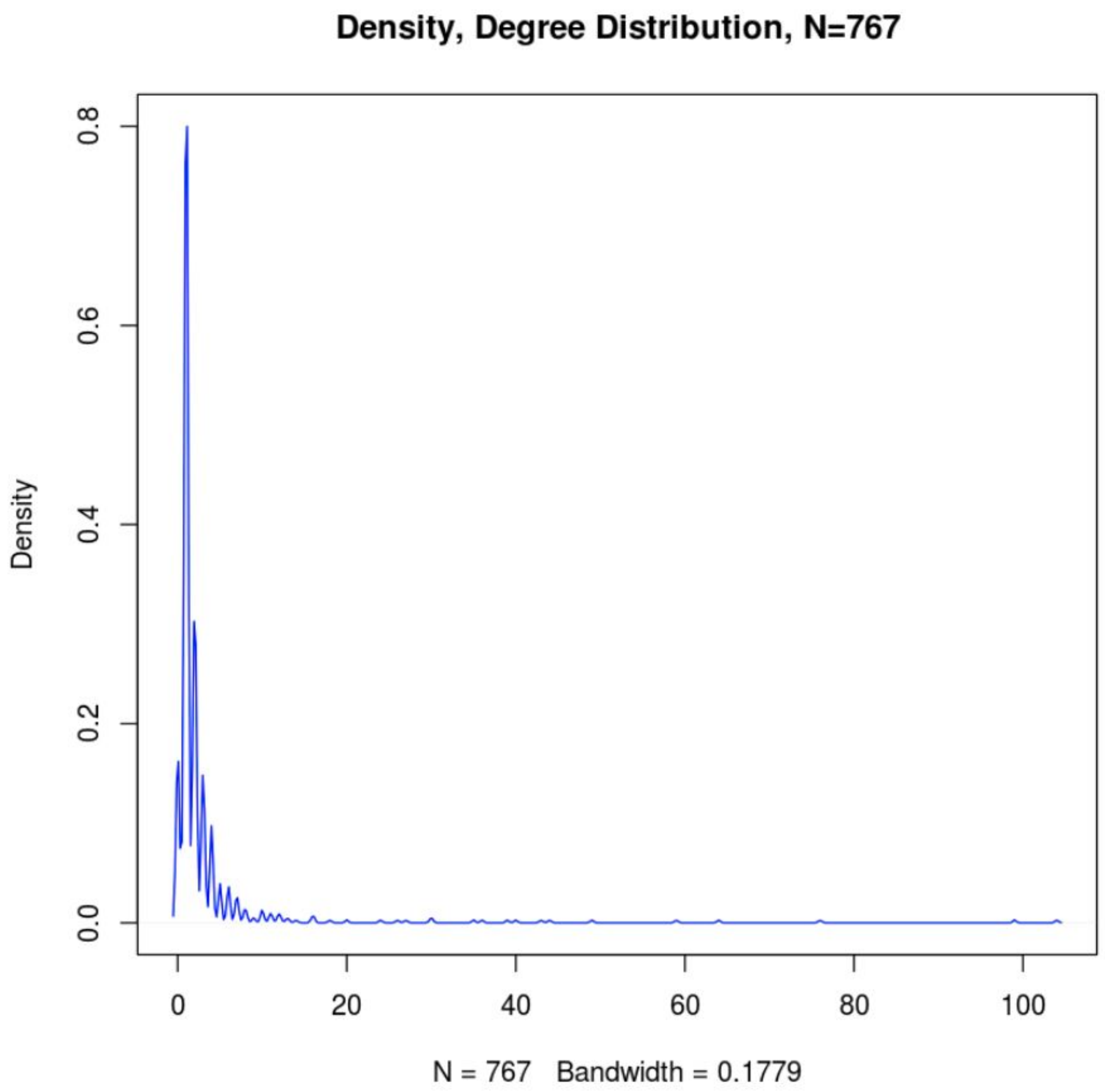


The above Fruchterman Reingold visualization colors the links between persons according to geographical region. Other categories were used for this coloring, but less visual correlation was represented. We expect that regional association will be a large factor in our model.

Next, we can look at the distribution of the edge weights (where parties mention each other in their messages). The the below plot, we can see that the distribution is skewed. Most of the values are less than five, and the maximum is 47.

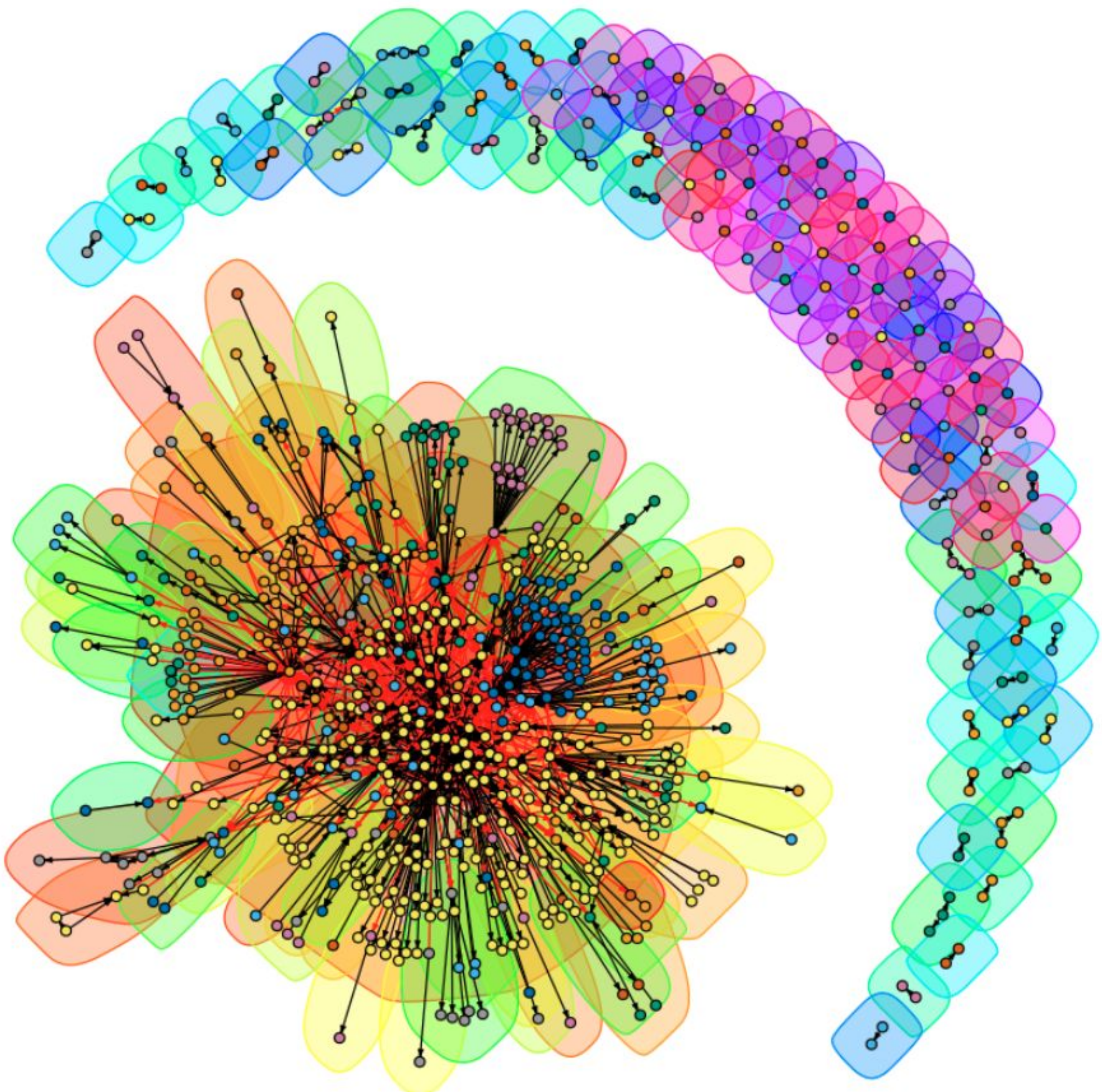


Next, we can consider the degree distribution. This tells us the probability distribution (fraction of the nodes in the network with degree k) based on the number of connection the node has to other nodes. Most nodes have less than five connections, but some have over 100!


Density, Degree Distribution, N=767
Density
0.0
0.2
0.4
0.6
0.8
0
20
40
60
80
100
N = 767 Bandwidth = 0.1779

Community Detection

The below community detection algorithm reprentended the strongest of a variety of algorithms. ‘Walktrap’ was better than ‘fast greedy’, ‘edge betweenness’, ‘spinglass’, ‘leading eigenvector’, and ‘propagation’.



This provides an interesting perspective as it shows what one may intuit about the distributions of active, public communications in an online course. Here we see that many students mention just one or two others. But a few dominant nodes, seemingly sharing a common geography, have many social mentions.

We can build two hypothesis based on community clustering. One is that many students pair up in small groups. These small communities are physically separate from the main network (on the upper right arc above). Recalling that node color is based on geographic region, we concluded a second hypothesis, based on many clusters (e.g. the blue, purple, and yellow in the center) that have obvious groupings in the main, central network.

Finally, we use mixing matrices to view the actual numerical relationships between our attributes discussed above.

```
mixingmatrix(net, "Continent")
```

```
                   To
From                Africa Asia Australia and NZ Europe International
  Africa                 0    1                0      1             0
  Asia                   0    3                2      3             0
  Australia and NZ       1    3               17     52             1
  Europe                 2   13               20     99             3
  International          0    0                0      0             0
  North America          1    6               13     51             3
  South America          0    2                0     20             0
  Unknown                0    3               14     50             3
  Total                  4   31               66    276            10
                   To
From                North America South America Unknown Total
  Africa                        0             0       0     2
  Asia                          4             1      11    24
  Australia and NZ             46            11      26   157
  Europe                      110            17      79   343
  International                 0             0       1     1
  North America               141            15      81   311
  South America                29            34      18   103
  Unknown                      73            19      90   252
  Total                       403            97     306  1193
```

These relationships are seemingly not as clear cut as the images portray. Additionally, over 25% of the nodes are unknown. Next, we look at the next 4 attributes:

```
mixingmatrix(net, "SocioTech")

           To
From        Social Technical Unknown Total
  Social      1078         2      49  1129
  Technical      0         0       0     0
  Unknown       21         0      43    64
  Total       1099         2      92  1193
```

```
mixingmatrix(net, "WorkType")

          To
From        Mixed Other Practice Research Unknown Total
  Mixed       51     0       60       21      11   143
  Other        1     0        1        0       2     4
  Practice   137     0      471       80      98   786
  Research    43     0       45       30      15   133
  Unknown     17     0       50        4      56   127
  Total      249     0      627      135     182  1193
```

```
# next explore the mixing matrix of the network
mixingmatrix(net, 'Role')

                    To
From                 Course Instructor Student Total
  Course Instructor                  1       9    10
  Student                           98    1085  1183
  Total                             99    1094  1193
```

```
mixingmatrix(net, "Gender")

         To
From        F   M Org Unknown Total
  F       143 244  67      30   484
  M       149 240  59      31   479
  Org      47  69  20      20   156
  Unknown   9  17   3      45    74
  Total   348 570 149     126  1193
```

Under the role above, it stands out that there is little public instructor interaction, and little instructor mentioning of students. Student's mainly mention students and about 8% of the time, instructors. In also looking at gender above, it appears as though male interaction is likely a factor we want want to consider.

Finally, we need to consider factors associated with the network's structural features. In the below summary table, we observe high two-path and intransitive properties.

```
# use ergm counts for elected structural features
summary(net ~ edges + mutual + triangles + simmelianties + intransitive + transitive + cyclicalties + twopath)

edges
      1193
mutual
      55
triangle
      1020
simmelianties
      18
intransitive
      7415
transitive
      964
cyclicalties
      108
twopath
      8379
```

Network Modeling

From the course notes we recall that goodness of fit is generated by the Monte Carlo Markov Chain algorithm in that it is used to approximate the Maximum Likelihood Estimation function.

Network modeling with ERGM starts with a null model that is used as a benchmark for other models to assess goodness of fit. Typically the null model isn't very good, since it only considers network density in relation to a randomly generated network. As we build in both network structural elements and certain edge and node attributes, the AIC coefficients tell us how the model is doing.

```r
## part III, modeling
null_model <- ergm(net ~ edges)
```

```
Evaluating log-likelihood at the estimate.
```

```r
summary(null_model)
```

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

Formula:   net ~ edges

Iterations:  8 out of 20

Monte Carlo MLE Results:
      Estimate Std. Error MCMC % p-value
edges -6.19741    0.02898      0  <1e-04 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

     Null Deviance: 814478  on 587522  degrees of freedom
 Residual Deviance:  17175  on 587521  degrees of freedom

AIC: 17177    BIC: 17189    (Smaller is better.)
```

```r
# find the z density
gden(net)
```

```
0.00203056225979623
```

Here we see that the network density is less than 50% which is not unusual. We also see that the zDensity is 0.00203. This means that the probability of creating an additional edge by adding one more node is very low.

```
summary(model_v1)

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

Formula:   net ~ edges + mutual + idegree(1) + odegree(1)

Iterations:  2 out of 20

Monte Carlo MLE Results:
         Estimate Std. Error MCMC % p-value
edges    -6.31164    0.03340      0 < 1e-04 ***
mutual    4.00271    0.14179      1 < 1e-04 ***
idegree1  0.04523    0.08071      0 0.57521
odegree1 -0.23137    0.08277      0 0.00519 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

     Null Deviance: 814478  on 587522  degrees of freedom
 Residual Deviance:  16858  on 587518  degrees of freedom

AIC: 16866    BIC: 16911    (Smaller is better.)
```

For my first of three non-baseline (null) models, I look at first order and mutual interactions. In the above summary, we can see that the mutual connections and out degree one, offer the strongest improvements over the null model (BIC improves from 17,189 to 16,911).

```
summary(model_v2)

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

Formula:   net ~ edges + mutual + gwidegree(0.6, fixed = T)

Iterations:  7 out of 20

Monte Carlo MLE Results:
          Estimate Std. Error MCMC % p-value
edges     -5.16274    0.03373      0  <1e-04 ***
mutual     4.03486    0.14796      1  <1e-04 ***
gwidegree -3.07379    0.09579      0  <1e-04 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

     Null Deviance: 814478  on 587522  degrees of freedom
 Residual Deviance:  15961  on 587519  degrees of freedom

AIC: 15967    BIC: 16001    (Smaller is better.)
```

For the second non-baseline model, I keep mutual and add gwidegree to control for degree and closure. According to the above summary, this offers further improvement from BIC 16,911 to 15,994.

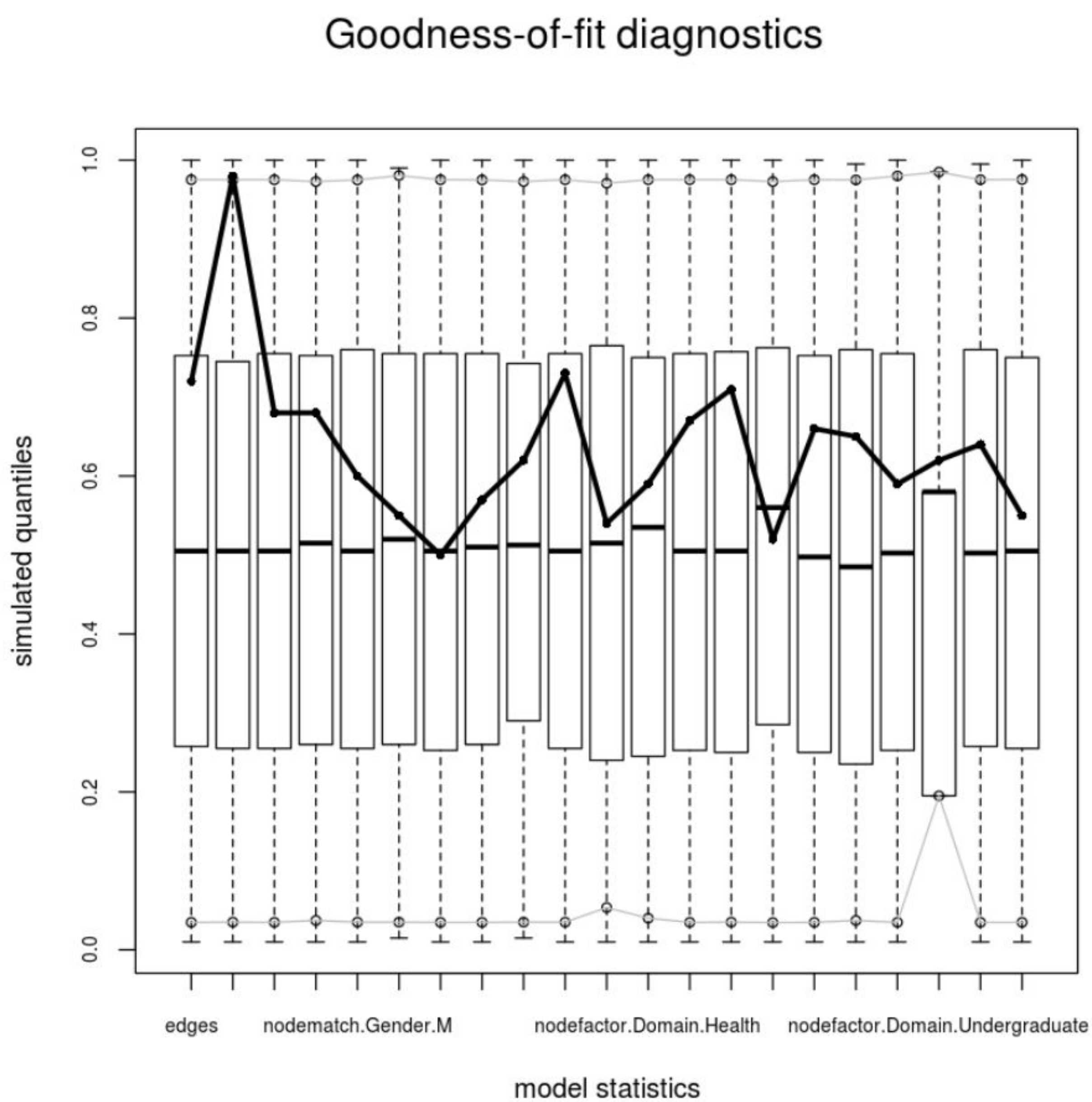
So, for the third model, what did we explore earlier that may help our model's performance? Domain and gender both looked interesting. In the below summary, we can see how these new additions impacted our model:

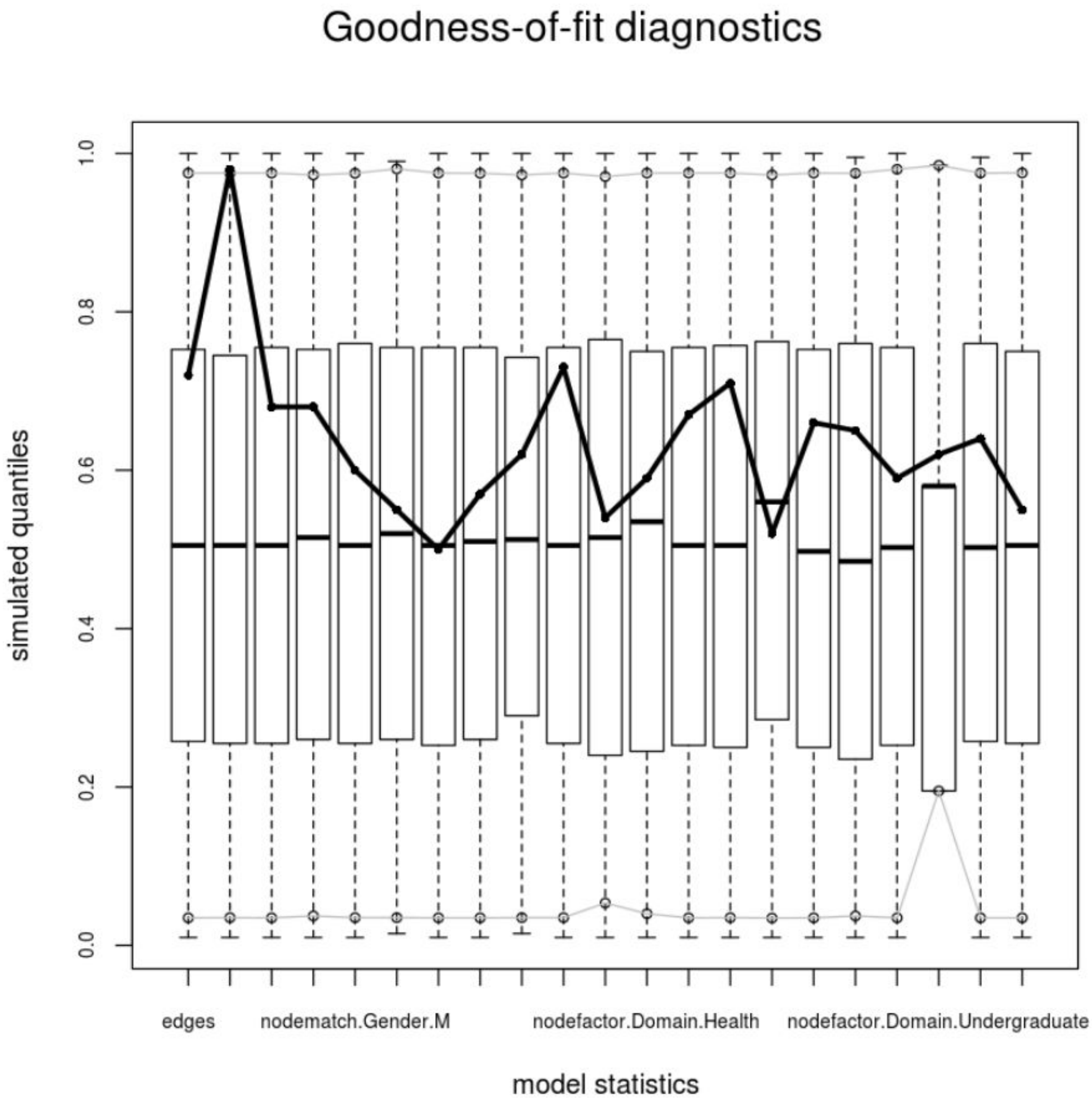
```
summary(model_v3)
```

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

Formula:   net ~ edges + mutual + gwidegree(0.6, fixed = T) + nodematch("Gender",
    diff = T) + nodefactor("Domain")

Iterations:  16 out of 20

Monte Carlo MLE Results:
                                                   Estimate Std. Error MCMC %
edges                                              -5.54253    0.12950      0
mutual                                              3.71491    0.16767      1
gwidegree                                          -2.75169    0.09909      0
nodematch.Gender.F                                  0.17227    0.08619      0
nodematch.Gender.M                                  0.19783    0.07221      0
nodematch.Gender.Org                               -0.26117    0.23549      0
nodematch.Gender.Unknown                            1.58809    0.18826      0
nodefactor.Domain.Community                         1.25815    0.15294      0
nodefactor.Domain.Elementary and primary education -0.14047    0.20760      0
nodefactor.Domain.Entrepreneurship                  0.34182    0.07483      0
nodefactor.Domain.Government                       -0.26587    0.34408      0
nodefactor.Domain.Health                            0.37759    0.12246      0
nodefactor.Domain.Higher Education                  0.23750    0.06679      0
nodefactor.Domain.Languages                         0.38243    0.07967      0
nodefactor.Domain.Library                          -0.10912    0.24161      0
nodefactor.Domain.Organization                      0.32405    0.08448      0
nodefactor.Domain.Other                            -0.32783    0.26564      0
nodefactor.Domain.Secondary education              -0.01437    0.08275      0
nodefactor.Domain.Undergraduate                    -1.82696    0.94976      0
nodefactor.Domain.Unknown                          -0.39930    0.08818      0
nodefactor.Domain.Various                           0.22400    0.09915      0
                                                    p-value
edges                                               < 1e-04 ***
mutual                                              < 1e-04 ***
gwidegree                                           < 1e-04 ***
nodematch.Gender.F                                 0.045651 *
nodematch.Gender.M                                 0.006149 **
nodematch.Gender.Org                               0.267420
nodematch.Gender.Unknown                            < 1e-04 ***
nodefactor.Domain.Community                         < 1e-04 ***
nodefactor.Domain.Elementary and primary education 0.498621
nodefactor.Domain.Entrepreneurship                  < 1e-04 ***
nodefactor.Domain.Government                       0.439698
nodefactor.Domain.Health                           0.002047 **
nodefactor.Domain.Higher Education                 0.000377 ***
nodefactor.Domain.Languages                         < 1e-04 ***
nodefactor.Domain.Library                          0.651526
nodefactor.Domain.Organization                     0.000125 ***
nodefactor.Domain.Other                            0.217172
nodefactor.Domain.Secondary education              0.862105
nodefactor.Domain.Undergraduate                    0.054404 .
nodefactor.Domain.Unknown                           < 1e-04 ***
nodefactor.Domain.Various                          0.023872 *
```

The BIC took a nice improvement (from 15,994 to 15,891).

The next plots are helpful in reviewing the model's goodness of fit, that is modelled versus observed performance. Here we run simulations and see that the model still struggles for some categories (while considering it it trying to capture edges, gender, and different domains):





Finally, we can observe the odds of the external and internal factors based on our initial thoughts and later findings:

```
# check the odds of the data
lapply(model_v3[1], exp)
```

$coef =

edges
: 0.00391661826485624

mutual
: 41.0547372320082

gwidegree
: 0.063819727764621

nodematch.Gender.F
: 1.18799674927322

nodematch.Gender.M
: 1.21875251802766

nodematch.Gender.Org
: 0.770151522368293

nodematch.Gender.Unknown
: 4.89437985237417

nodefactor.Domain.Community
: 3.51890637065117

nodefactor.Domain.Elementary and primary education
: 0.868945447973084

nodefactor.Domain.Entrepreneurship
: 1.40750648049384

nodefactor.Domain.Government
: 0.766535273380976

nodefactor.Domain.Health
: 1.45876545194954

nodefactor.Domain.Higher Education
: 1.26808114702961

nodefactor.Domain.Languages
: 1.4658360717404

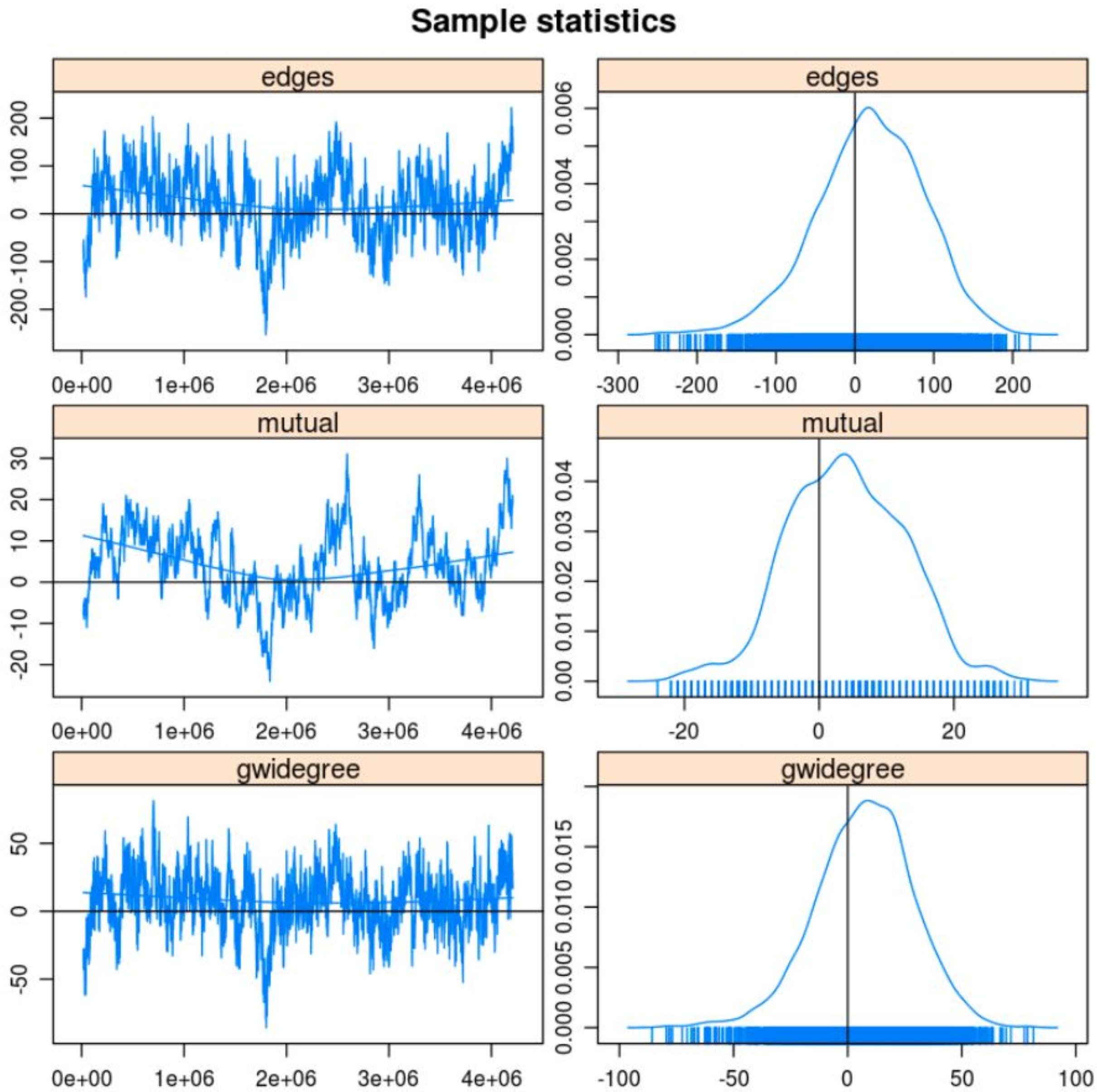
nodefactor.Domain.Library
: 0.896620272033604

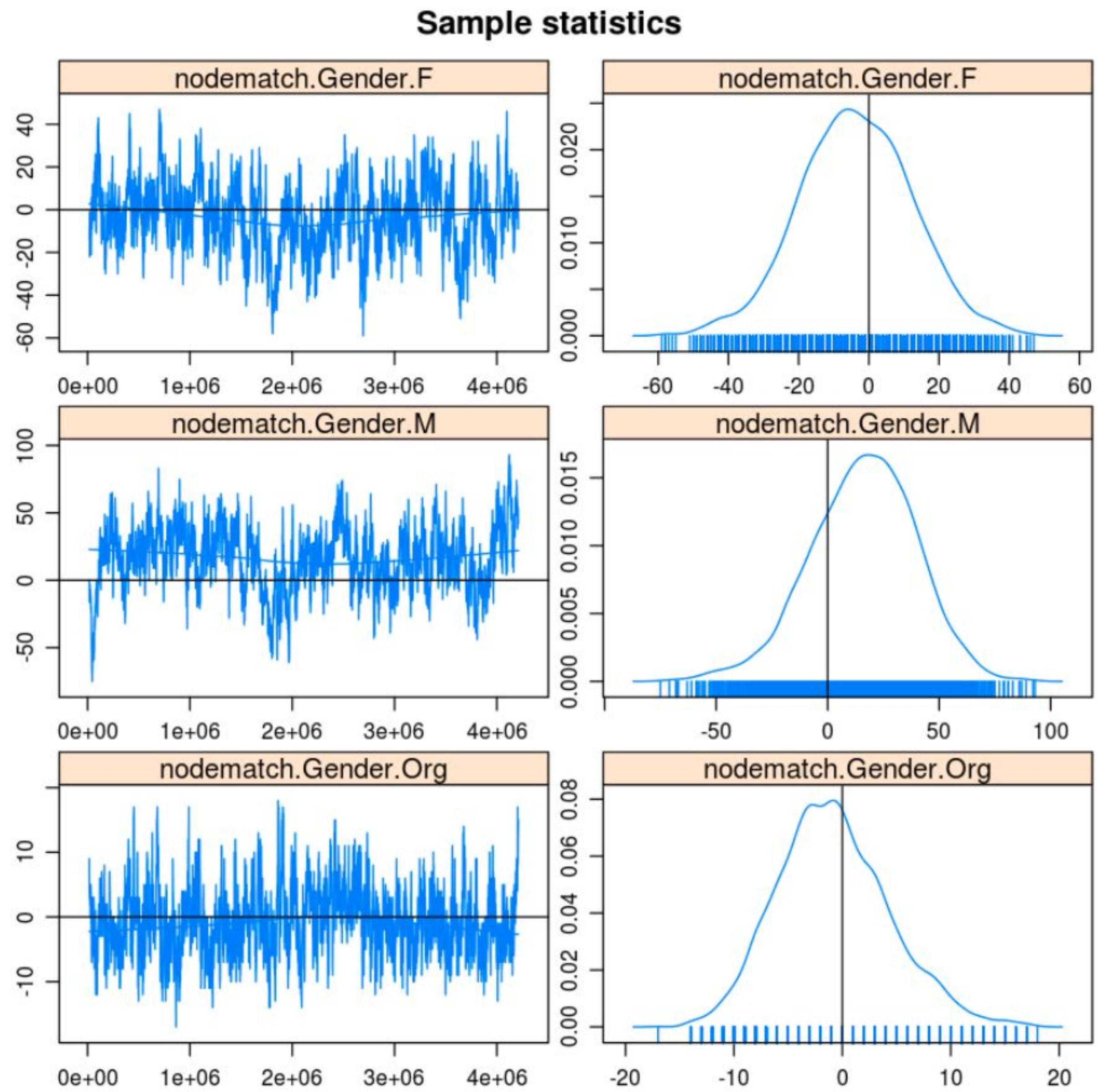
nodefactor.Domain.Organization

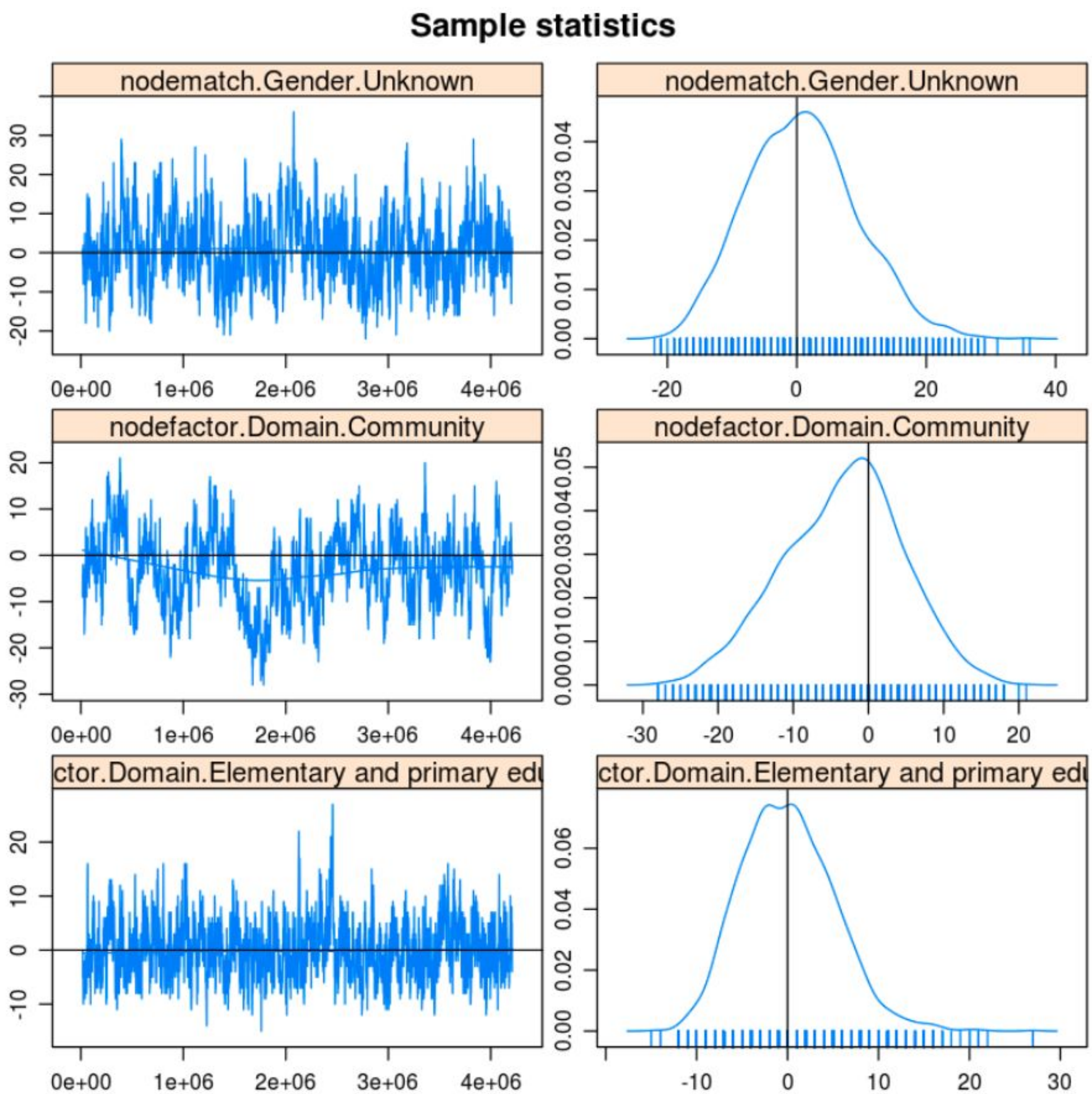
Based on the above, we can observe some homophily between males, higher education, institution accounts, community members, and entrepreneurs. These subgroups were factors of the information network's flow and structural reciprocity.

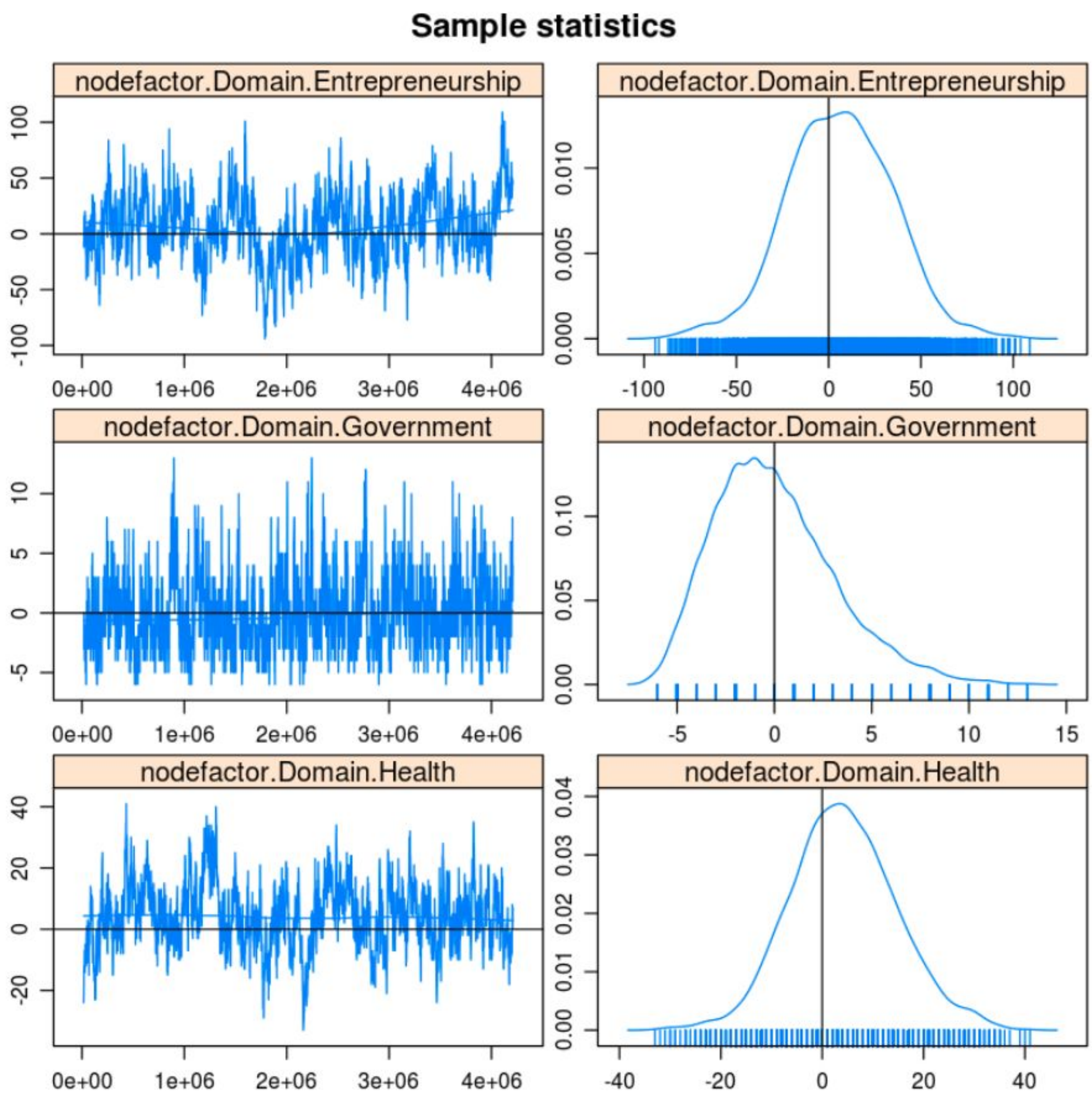
This analysis is publically available on my GitHub account under the creative commons license as it may be helpful for others studying social information exchanges or comparison between large online program behaviors.

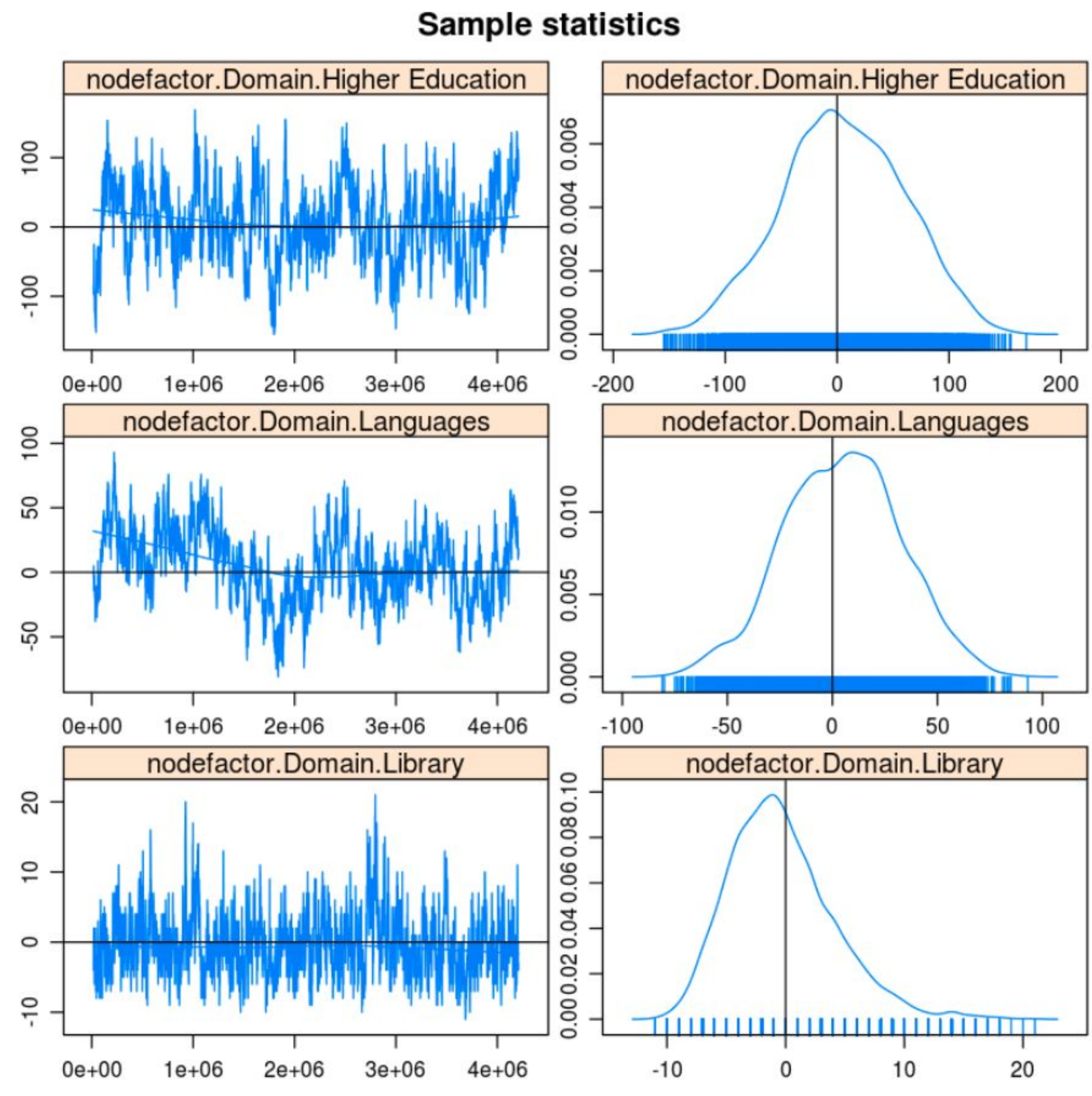
Appendix A : Model III Sample Statistics Plots

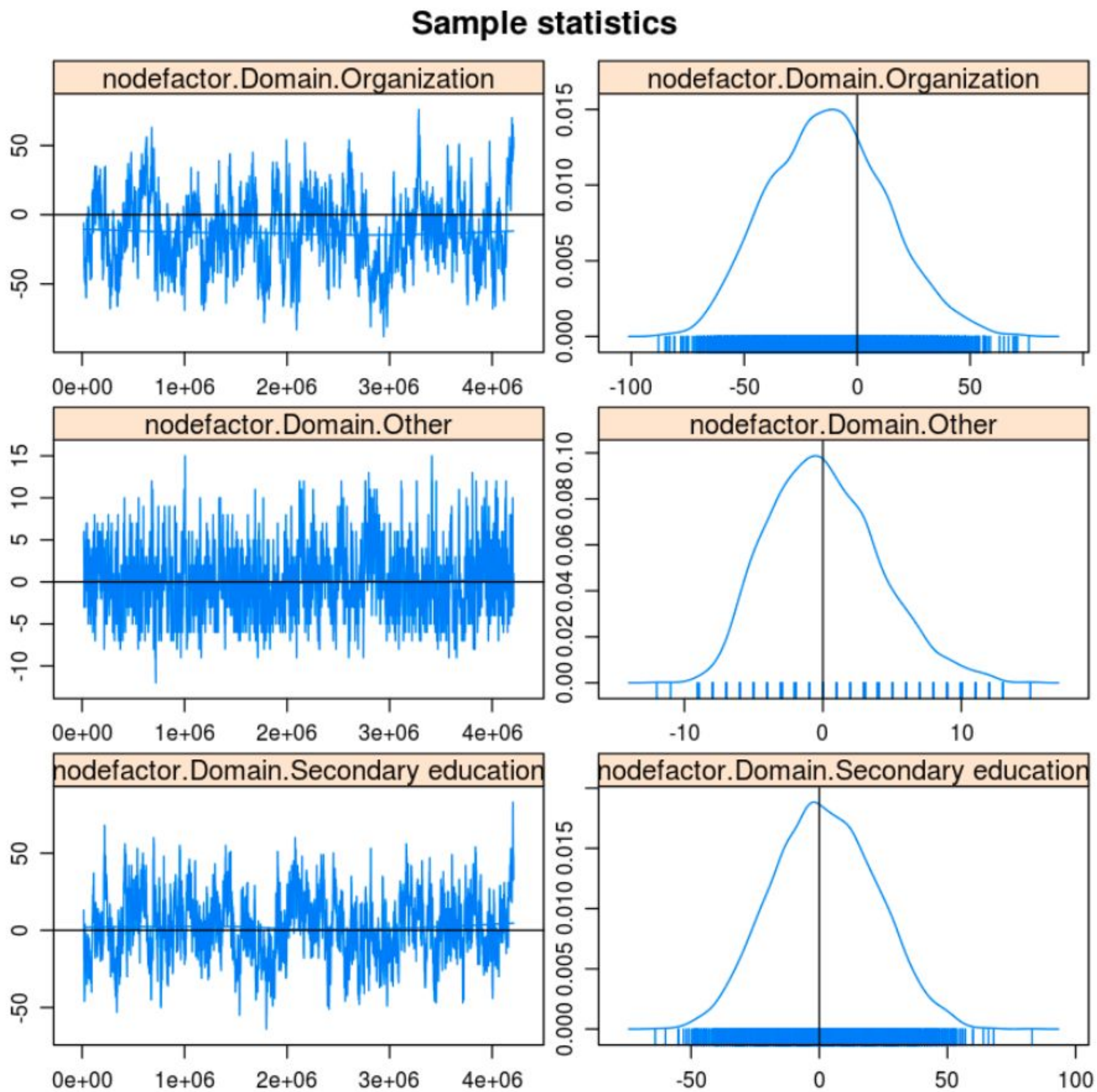


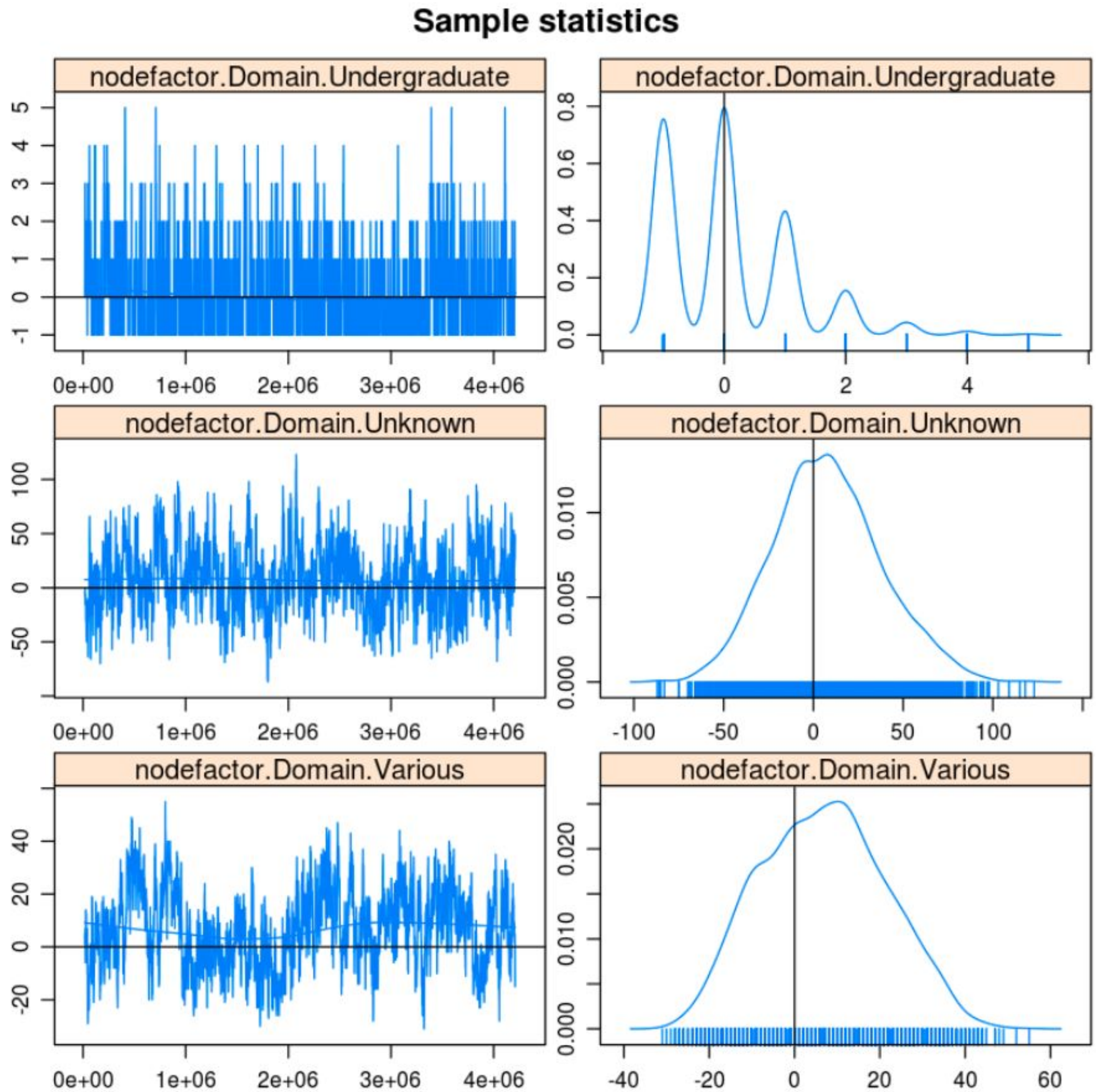

Sample statistics
nodematch.Gender.F
nodematch.Gender.M
nodematch.Gender.Org


Sample statistics
nodematch.Gender.Unknown
nodematch.Gender.Unknown
nodefactor.Domain.Community
nodefactor.Domain.Community
ctor.Domain.Elementary and primary edu
ctor.Domain.Elementary and primary edu
0e+00
1e+06
2e+06
3e+06
4e+06


Sample statistics
nodefactor.Domain.Entrepreneurship
nodefactor.Domain.Entrepreneurship
nodefactor.Domain.Government
nodefactor.Domain.Government
nodefactor.Domain.Health
nodefactor.Domain.Health


Sample statistics
nodefactor.Domain.Higher Education
nodefactor.Domain.Higher Education
nodefactor.Domain.Languages
nodefactor.Domain.Languages
nodefactor.Domain.Library
nodefactor.Domain.Library


Sample statistics
nodefactor.Domain.Organization
nodefactor.Domain.Organization
nodefactor.Domain.Other
nodefactor.Domain.Other
nodefactor.Domain.Secondary education
nodefactor.Domain.Secondary education
0e+00
1e+06
2e+06
3e+06
4e+06


Sample statistics
nodefactor.Domain.Undergraduate
nodefactor.Domain.Undergraduate
nodefactor.Domain.Unknown
nodefactor.Domain.Unknown
nodefactor.Domain.Various
nodefactor.Domain.Various
0e+00
1e+06
2e+06
3e+06
4e+06

R Code

```
library(igraph)
```

```
Attaching package: 'igraph'

The following objects are masked from 'package:stats':

    decompose, spectrum

The following object is masked from 'package:base':

    union
```

```
setwd('/home/cbios/github/swib_data_quality_analysis/EDMTechnicalTasks/application_analysis/week10')
```

```
g <- read_graph("cck11_dataset.gml", format=c("gml"))
```

```
summary(g)
```

```
IGRAPH 60db74a DNW- 767 1193 --
+ attr: id (v/n), name (v/c), Role (v/c), SocioTech (v/c), Gender
| (v/c), Domain (v/c), WorkType (v/c), Continent (v/c), weight (e/n)
```

```
# extract the dataframes with node and edge information
g_net.v.df <- as_data_frame(g, what='vertices')
g_net.e.df <- as_data_frame(g, what='edges')
```

```
# write the split network to csv
write.csv(g_net.v.df, file="cmooc_v.csv")
write.csv(g_net.e.df, file="cmooc_e.csv")
```

```
# also export to gexf for networks analysis using python
graph_gexf <- igraph.to.gexf(g, position=NULL)
```

```
Error in igraph.to.gexf(g, position = NULL): could not find function "igraph.to.gexf"
Traceback:
```

```
# take a look at the communities:
wc <- walktrap.community(g)
modularity(wc)
#wc <- fastgreedy.community(g)
#modularity(wc)
#wc <- edge.betweenness.community(g)
#modularity(wc)
#wc <- spinglass.community(g)
#modularity(wc)
#wc <- leading.eigenvector.community(g)
#modularity(wc)
#wc <- label.propagation.community(g)
#modularity(wc)

# Plot
par(mfrow=c(1,1), mar=rep(1,4))
layout <-layout.fruchterman.reingold(g)
plot(wc, g, layout=layout, vertex.label=NA, vertex.size=2,  edge.arrow.size=.2)
```

0.502232670783997

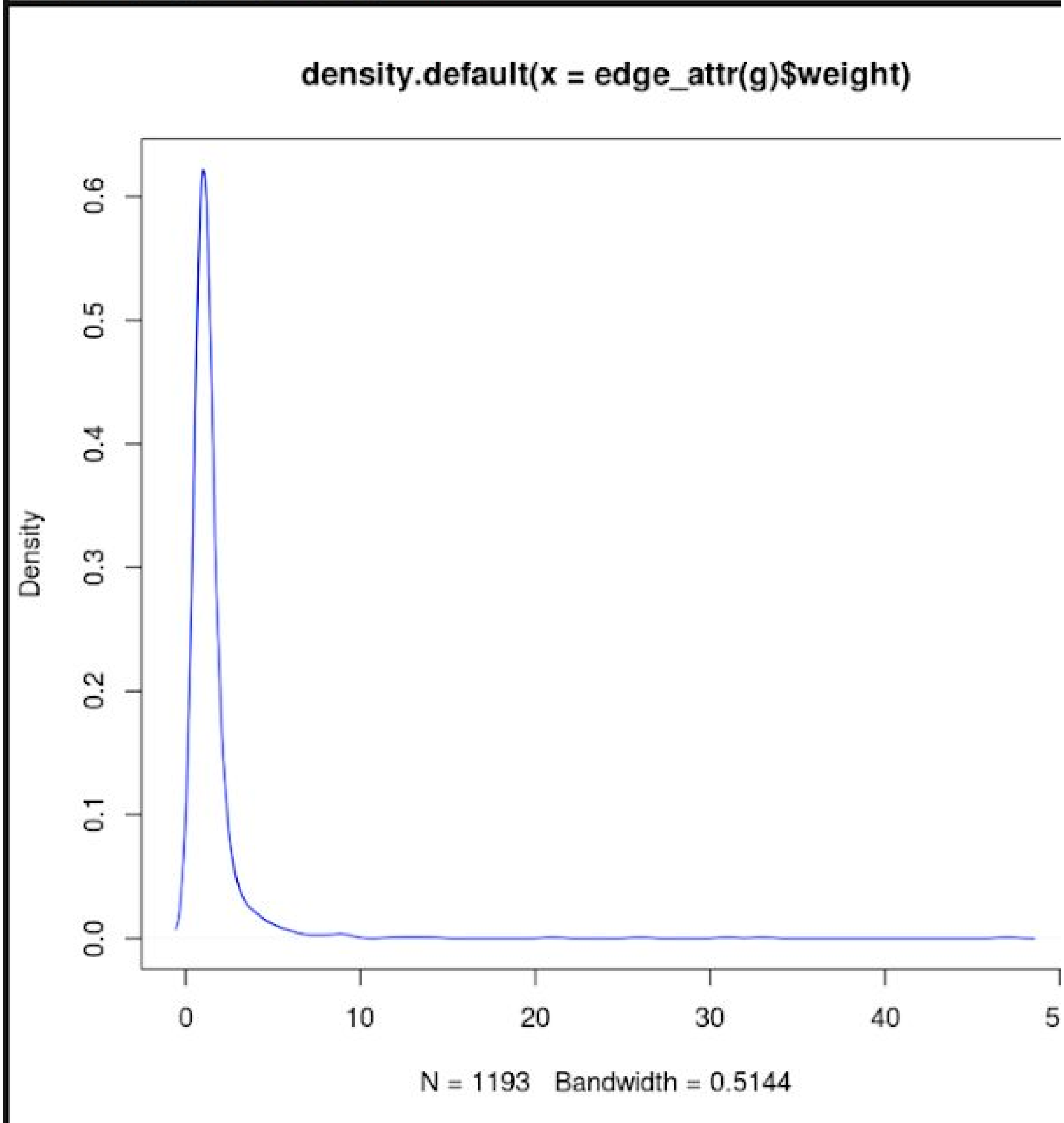
```
table(V(g)$Gender)
```

```
     F       M     Org Unknown
   237     285     100     145
```

```
table(V(g)$WorkType)
```

```
   Mixed    Other Practice Research  Unknown
      66        2      400      107      192
```

```
plot(density(edge_attr(g)$weight), col='blue')
```




```
summary(E(g)$weight)
```

```
# ergm needs iGraph, convert
net <- asNetwork(g)
```

```
net
```

```
 Network attributes:
  vertices = 767
  directed = TRUE
  hyper = FALSE
  loops = FALSE
  multiple = FALSE
  bipartite = FALSE
  total edges= 1193
    missing edges= 0
    non-missing edges= 1193

 Vertex attribute names:
    Continent Domain Gender id Role SocioTech vertex.names WorkType

 Edge attribute names not shown
```

```
list.vertex.attributes(net)
```

1. 'Continent'
2. 'Domain'
3. 'Gender'
4. 'id'
5. 'na'
6. 'Role'
7. 'SocioTech'
8. 'vertex.names'
9. 'WorkType'

```
mixingmatrix(net, "SocioTech")
```

```
           To
From        Social Technical Unknown Total
  Social      1078         2      49  1129
  Technical      0         0       0     0
  Unknown       21         0      43    64
  Total       1099         2      92  1193
```

```
mixingmatrix(net, "WorkType")
```

```
          To
From       Mixed Other Practice Research Unknown Total
  Mixed       51     0       60       21      11   143
  Other        1     0        1        0       2     4
  Practice   137     0      471       80      98   786
  Research    43     0       45       30      15   133
  Unknown     17     0       58        4      56   127
  Total      249     0      627      135     182  1193
```

```
# next explore the mixing matrix of the network
mixingmatrix(net, 'Role')
```

```
                  To
From                Course Instructor Student Total
  Course Instructor                 1       9    10
  Student                          98    1085  1183
  Total                            99    1094  1193
```

```
                                     To
From                                  Higher Education Languages Library
  Business                                          18         8       2
  Community                                         19         3       0
  Elementary and primary education                   2         0       0
  Entrepreneurship                                  55         8       0
  Government                                         0         0       0
  Health                                             8         1       0
  Higher Education                                 164        20       1
  Languages                                         28        26       1
  Library                                            5         0       0
  Organization                                      39         4       1
  Other                                              1         0       0
  Secondary education                               19        16       0
  Undergraduate                                      0         1       0
  Unknown                                           30        11       1
  Various                                           17         2       0
  Total                                            405       100       6
                                     To
From                                  Organization Other Secondary education
  Business                                      10     0                   4
  Community                                      5     2                   3
  Elementary and primary education               3     0                   5
  Entrepreneurship                              16     1                  16
  Government                                     0     0                   0
  Health                                         2     0                   0
  Higher Education                              28     3                   8
  Languages                                      2     0                   4
  Library                                        0     0                   0
  Organization                                  17     1                  11
  Other                                          0     0                   0
  Secondary education                           18     2                  15
  Undergraduate                                  0     0                   0
  Unknown                                       16     0                   2
  Various                                        1     0                   2
  Total                                        118     9                  70
                                     To
From                                  Undergraduate Unknown Various Total
  Business                                        0      10       3    67
  Community                                       0      10       2    57
  Elementary and primary education                0       0       0    12
  Entrepreneurship                                0      21       5   164
  Government                                      0       0       0     1
  Health                                          0       2       1    20
  Higher Education                                0      39      17   338
  Languages                                       0      12       3   100
  Library                                         0       0       0     5
  Organization                                    0      30       3   131
  Other                                           0       1       0     3
  Secondary education                             0      11       1   109
  Undergraduate                                   0       0       0     1
  Unknown                                         0      58       8   151
  Various                                         0       4       1    34
  Total                                           0     198      44  1193
```

```
mixingmatrix(net, "Continent")
```

```
                   To
From                 Africa Asia Australia and NZ Europe International
  Africa                  0    1                0      1             0
  Asia                    0    3                2      3             0
  Australia and NZ        1    3               17     52             1
  Europe                  2   13               20     99             3
  International           0    0                0      0             0
  North America           1    6               13     51             3
  South America           0    2                0     20             0
  Unknown                 0    3               14     50             3
  Total                   4   31               66    276            10
                   To
From                 North America South America Unknown Total
  Africa                         0             0       0     2
  Asia                           4             1      11    24
  Australia and NZ              46            11      26   157
  Europe                       110            17      79   343
  International                  0             0       1     1
```

```
nodefactor.Continent.Asia
        55
nodefactor.Continent.Australia and NZ
        223
nodefactor.Continent.Europe
        619
nodefactor.Continent.International
        11
nodefactor.Continent.North America
        714
nodefactor.Continent.South America
        200
nodefactor.Continent.Unknown
        558
nodematch.Continent.Africa
        0
nodematch.Continent.Asia
        3
nodematch.Continent.Australia and NZ
        17
nodematch.Continent.Europe
        99
nodematch.Continent.International
        0
nodematch.Continent.North America
        141
nodematch.Continent.South America
        34
nodematch.Continent.Unknown
        90
mix.Continent.Africa.Africa
        0
mix.Continent.Asia.Africa
        0
mix.Continent.Australia and NZ.Africa
        1
mix.Continent.Europe.Africa
        2
mix.Continent.International.Africa
        0
mix.Continent.North America.Africa
        1
mix.Continent.South America.Africa
        0
mix.Continent.Unknown.Africa
        0
mix.Continent.Africa.Asia
        1
mix.Continent.Asia.Asia
        3
```

```
mix.Continent.Africa.North America
        0
mix.Continent.Asia.North America
        4
mix.Continent.Australia and NZ.North America
        46
mix.Continent.Europe.North America
        110
mix.Continent.International.North America
        0
mix.Continent.North America.North America
        141
mix.Continent.South America.North America
        29
mix.Continent.Unknown.North America
        73
mix.Continent.Africa.South America
        0
mix.Continent.Asia.South America
        1
mix.Continent.Australia and NZ.South America
        11
mix.Continent.Europe.South America
        17
mix.Continent.International.South America
        0
mix.Continent.North America.South America
        15
mix.Continent.South America.South America
        34
mix.Continent.Unknown.South America
        19
mix.Continent.Africa.Unknown
        0
mix.Continent.Asia.Unknown
        11
mix.Continent.Australia and NZ.Unknown
        26
mix.Continent.Europe.Unknown
        79
mix.Continent.International.Unknown
        1
mix.Continent.North America.Unknown
        81
mix.Continent.South America.Unknown
        18
mix.Continent.Unknown.Unknown
        90
```

```
## part III, modeling
null_model <- ergm(net ~ edges)
```

```
Evaluating log-likelihood at the estimate.
```

```
summary(null_model)
```

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

```
summary(model_v2)
```

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

Formula:   net ~ edges + mutual + gwidegree(0.6, fixed = T)

Iterations:  7 out of 20

Monte Carlo MLE Results:
          Estimate Std. Error MCMC % p-value
edges     -5.16442    0.03377      0  <1e-04 ***
mutual     4.04740    0.16913      1  <1e-04 ***
gwidegree -3.07472    0.09431      0  <1e-04 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

     Null Deviance: 814478  on 587522  degrees of freedom
 Residual Deviance:  15954  on 587519  degrees of freedom

AIC: 15960    BIC: 15994    (Smaller is better.)
```

```
#^^ not much improvement via structural components, lets look at attributes

model_v3 <- ergm( net ~ edges + mutual + gwidegree(0.6, fixed=T) + nodematch("Gender", diff=T) + nodefactor("Domain") )
```

```
summary(model_v3)
```

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

Formula:   net ~ edges + mutual + gwidegree(0.6, fixed = T) + nodematch("Gender",
    diff = T) + nodefactor("Domain")

Iterations:  16 out of 20

Monte Carlo MLE Results:
                                                   Estimate Std. Error MCMC %
edges                                              -5.54253    0.12950      0
mutual                                              3.71491    0.16767      1
gwidegree                                          -2.75169    0.09909      0
nodematch.Gender.F                                  0.17227    0.08619      0
nodematch.Gender.M                                  0.19783    0.07221      0
nodematch.Gender.Org                               -0.26117    0.23549      0
nodematch.Gender.Unknown                            1.58809    0.18826      0
nodefactor.Domain.Community                         1.25815    0.15294      0
nodefactor.Domain.Elementary and primary education -0.14047    0.20760      0
nodefactor.Domain.Entrepreneurship                  0.34182    0.07483      0
nodefactor.Domain.Government                       -0.26587    0.34408      0
nodefactor.Domain.Health                            0.37759    0.12246      0
nodefactor.Domain.Higher Education                  0.23750    0.06679      0
nodefactor.Domain.Languages                         0.38243    0.07967      0
nodefactor.Domain.Library                          -0.10912    0.24161      0
nodefactor.Domain.Organization                      0.32405    0.08448      0
nodefactor.Domain.Other                            -0.32783    0.26564      0
nodefactor.Domain.Secondary education              -0.01437    0.08275      0
nodefactor.Domain.Undergraduate                    -1.82696    0.94976      0
nodefactor.Domain.Unknown                          -0.39930    0.08818      0
nodefactor.Domain.Various                           0.22400    0.09915      0
                                                    p-value
edges                                               < 1e-04 ***
mutual                                              < 1e-04 ***
gwidegree                                           < 1e-04 ***
nodematch.Gender.F                                 0.045651 *
nodematch.Gender.M                                 0.006149 **
nodematch.Gender.Org                               0.267420
nodematch.Gender.Unknown                            < 1e-04 ***
nodefactor.Domain.Community                         < 1e-04 ***
nodefactor.Domain.Elementary and primary education 0.498621
nodefactor.Domain.Entrepreneurship                  < 1e-04 ***
nodefactor.Domain.Government                       0.439698
nodefactor.Domain.Health                           0.002047 **
nodefactor.Domain.Higher Education                 0.000377 ***
nodefactor.Domain.Languages                         < 1e-04 ***
nodefactor.Domain.Library                          0.651526
nodefactor.Domain.Organization                     0.000125 ***
nodefactor.Domain.Other                            0.217172
```

```
mcmc.diagnostics(model_v3)

Sample statistics summary:

Iterations = 16384:4209664
Thinning interval = 1024
Number of chains = 1
Sample size per chain = 4096

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

                                                  Mean     SD Naive SE
edges                                        19.266846 68.417  1.06901
mutual                                        4.069824  8.574  0.13397
gwidegree                                     7.252364 21.571  0.33705
nodematch.Gender.F                           -3.675781 16.091  0.25142
nodematch.Gender.M                           15.806885 23.821  0.37221
nodematch.Gender.Org                         -0.817871  5.196  0.08119
nodematch.Gender.Unknown                      0.860596  8.484  0.13257
nodefactor.Domain.Community                  -3.378418  7.961  0.12439
nodefactor.Domain.Elementary and primary education  0.063232  5.141  0.08033
nodefactor.Domain.Entrepreneurship            6.082031 28.929  0.45202
nodefactor.Domain.Government                 -0.008057  3.079  0.04810
nodefactor.Domain.Health                      4.220703 10.528  0.16450
nodefactor.Domain.Higher Education            5.509521 54.677  0.85433
nodefactor.Domain.Languages                   4.876709 27.713  0.43301
nodefactor.Domain.Library                    -0.335449  4.450  0.06953
nodefactor.Domain.Organization              -12.954834 25.722  0.40191
nodefactor.Domain.Other                       0.088623  3.958  0.06184
nodefactor.Domain.Secondary education         1.992676 20.216  0.31587
nodefactor.Domain.Undergraduate               0.086182  1.062  0.01659
nodefactor.Domain.Unknown                     7.899414 30.145  0.47101
nodefactor.Domain.Various                     6.755859 14.572  0.22769
                                             Time-series SE
edges                                               7.58354
mutual                                              1.68879
gwidegree                                           2.15759
nodematch.Gender.F                                  1.42752
nodematch.Gender.M                                  2.57811
nodematch.Gender.Org                                0.33354
nodematch.Gender.Unknown                            0.68204
nodefactor.Domain.Community                         0.92044
nodefactor.Domain.Elementary and primary education  0.27190
nodefactor.Domain.Entrepreneurship                  2.90808
nodefactor.Domain.Government                        0.16688
nodefactor.Domain.Health                            0.98073
nodefactor.Domain.Higher Education                  5.74094
nodefactor.Domain.Languages                         3.54856
nodefactor.Domain.Library                           0.30089
nodefactor.Domain.Organization                      2.37463
nodefactor.Domain.Other                             0.17827
nodefactor.Domain.Secondary education               1.65031
nodefactor.Domain.Undergraduate                     0.03619
nodefactor.Domain.Unknown                           2.16518
nodefactor.Domain.Various                           1.40319

2. Quantiles for each variable:

                                                2.5%     25%     50%
edges                                        -124.62 -24.000  21.000
mutual                                        -13.00  -2.000   4.000
gwidegree                                     -38.67  -6.095   8.025
nodematch.Gender.F                            -36.00 -14.250  -4.000
nodematch.Gender.M                            -34.00   0.000  17.000
```

```
                                                   gwidegree
edges                                             0.86008538
mutual                                            0.51651260
gwidegree                                         1.00000000
nodematch.Gender.F                                0.45301735
nodematch.Gender.M                                0.55507574
nodematch.Gender.Org                              0.17871492
nodematch.Gender.Unknown                          0.20313068
nodefactor.Domain.Community                       0.17706779
nodefactor.Domain.Elementary and primary education 0.10590116
nodefactor.Domain.Entrepreneurship                0.41618427
nodefactor.Domain.Government                      0.09402141
nodefactor.Domain.Health                          0.24132331
nodefactor.Domain.Higher Education                0.67717052
nodefactor.Domain.Languages                       0.48205690
nodefactor.Domain.Library                         0.17034845
nodefactor.Domain.Organization                    0.41154703
nodefactor.Domain.Other                           0.13379807
nodefactor.Domain.Secondary education             0.51759758
nodefactor.Domain.Undergraduate                   0.02488983
nodefactor.Domain.Unknown                         0.53980556
nodefactor.Domain.Various                         0.27038192
                                                   nodematch.Gender.F
edges                                                     0.528354614
mutual                                                    0.385362502
gwidegree                                                 0.453017353
nodematch.Gender.F                                        1.000000000
nodematch.Gender.M                                        0.194959927
nodematch.Gender.Org                                     -0.032426972
nodematch.Gender.Unknown                                  0.069254999
nodefactor.Domain.Community                               0.216467501
nodefactor.Domain.Elementary and primary education        0.027882659
nodefactor.Domain.Entrepreneurship                        0.370144093
nodefactor.Domain.Government                              0.060540386
nodefactor.Domain.Health                                  0.217866345
nodefactor.Domain.Higher Education                        0.471294761
nodefactor.Domain.Languages                               0.422718606
nodefactor.Domain.Library                                 0.130870933
nodefactor.Domain.Organization                            0.059427560
nodefactor.Domain.Other                                   0.090263623
nodefactor.Domain.Secondary education                     0.276771427
nodefactor.Domain.Undergraduate                           0.007284831
nodefactor.Domain.Unknown                                 0.249892747
nodefactor.Domain.Various                                 0.088739908
                                                   nodematch.Gender.M
edges                                                      0.68153433
mutual                                                     0.48315349
gwidegree                                                  0.55507574
nodematch.Gender.F                                         0.19495993
nodematch.Gender.M                                         1.00000000
nodematch.Gender.Org                                       0.02691665
nodematch.Gender.Unknown                                  -0.06224170
nodefactor.Domain.Community                                0.21886666
nodefactor.Domain.Elementary and primary education         0.14333064
nodefactor.Domain.Entrepreneurship                         0.49555856
nodefactor.Domain.Government                               0.10231918
nodefactor.Domain.Health                                   0.16632984
nodefactor.Domain.Higher Education                         0.57067780
nodefactor.Domain.Languages                                0.42854958
nodefactor.Domain.Library                                  0.09293787
nodefactor.Domain.Organization                             0.18268736
nodefactor.Domain.Other                                    0.08270728
nodefactor.Domain.Secondary education                      0.35579405
nodefactor.Domain.Undergraduate                            0.04906837
nodefactor.Domain.Unknown                                  0.15820592
nodefactor.Domain.Various                                  0.34371169
```

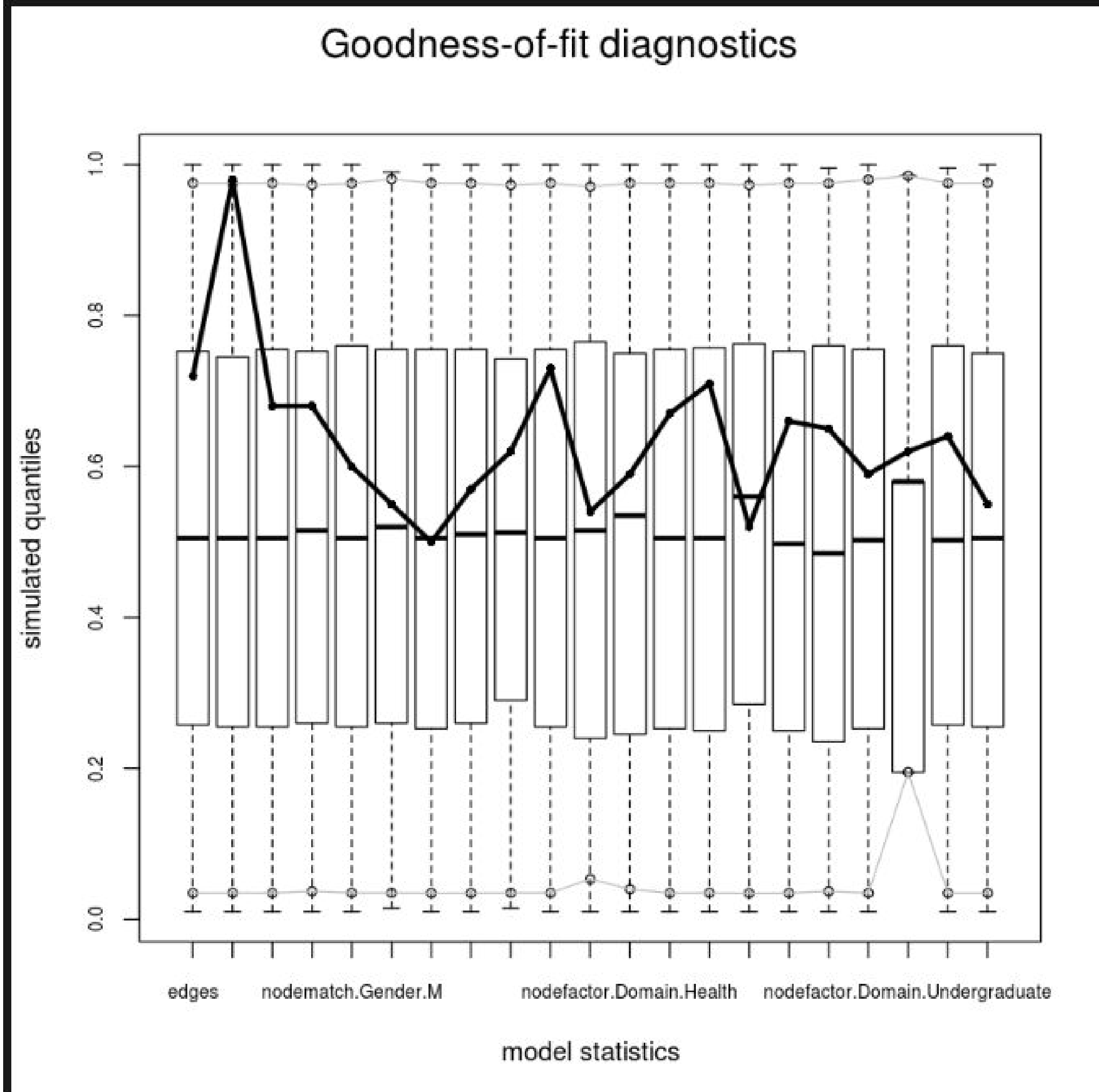
```
                                            nodefactor.Domain.Elementary and primary education
edges                                                                              0.141113769
mutual                                                                             0.102150121
gwidegree                                                                          0.105901161
nodematch.Gender.F                                                                 0.027882659
nodematch.Gender.M                                                                 0.143330641
nodematch.Gender.Org                                                               0.000473779
nodematch.Gender.Unknown                                                          -0.019639454
nodefactor.Domain.Community                                                        0.136339281
nodefactor.Domain.Elementary and primary education                                 1.000000000
nodefactor.Domain.Entrepreneurship                                                 0.033433750
nodefactor.Domain.Government                                                       0.055039018
nodefactor.Domain.Health                                                           0.059992529
nodefactor.Domain.Higher Education                                                 0.058865493
nodefactor.Domain.Languages                                                        0.086276039
nodefactor.Domain.Library                                                          0.026556163
nodefactor.Domain.Organization                                                     0.032762692
nodefactor.Domain.Other                                                           -0.024988687
nodefactor.Domain.Secondary education                                              0.064805246
nodefactor.Domain.Undergraduate                                                    0.083976340
nodefactor.Domain.Unknown                                                          0.048084921
nodefactor.Domain.Various                                                          0.062141859
                                            nodefactor.Domain.Entrepreneurship
edges                                                                0.57091147
mutual                                                               0.42213042
gwidegree                                                            0.41618427
nodematch.Gender.F                                                   0.37014409
nodematch.Gender.M                                                   0.49555856
nodematch.Gender.Org                                                 0.08978220
nodematch.Gender.Unknown                                            -0.03036702
nodefactor.Domain.Community                                          0.16718569
nodefactor.Domain.Elementary and primary education                   0.03343375
nodefactor.Domain.Entrepreneurship                                   1.00000000
nodefactor.Domain.Government                                        -0.01662256
nodefactor.Domain.Health                                             0.14728925
nodefactor.Domain.Higher Education                                   0.29598120
nodefactor.Domain.Languages                                          0.26953581
nodefactor.Domain.Library                                            0.09568942
nodefactor.Domain.Organization                                       0.18112091
nodefactor.Domain.Other                                              0.10556981
nodefactor.Domain.Secondary education                                0.19929528
nodefactor.Domain.Undergraduate                                      0.01749452
nodefactor.Domain.Unknown                                            0.18156115
nodefactor.Domain.Various                                            0.22135090
                                            nodefactor.Domain.Government
edges                                                      0.1060584973
mutual                                                    -0.0248457125
gwidegree                                                  0.0940214063
nodematch.Gender.F                                         0.0605403864
nodematch.Gender.M                                         0.1023191830
nodematch.Gender.Org                                      -0.0139978644
nodematch.Gender.Unknown                                   0.1072483065
nodefactor.Domain.Community                                0.1270138740
nodefactor.Domain.Elementary and primary education         0.0550390183
nodefactor.Domain.Entrepreneurship                        -0.0166225618
nodefactor.Domain.Government                               1.0000000000
nodefactor.Domain.Health                                   0.0306891265
nodefactor.Domain.Higher Education                         0.0505915646
nodefactor.Domain.Languages                                0.0260785379
nodefactor.Domain.Library                                 -0.0449377846
nodefactor.Domain.Organization                             0.0119265399
nodefactor.Domain.Other                                   -0.0357382492
nodefactor.Domain.Secondary education                      0.0729310217
nodefactor.Domain.Undergraduate                           -0.0003852756
nodefactor.Domain.Unknown                                  0.1303084330
nodefactor.Domain.Various                                  0.0271954551
                                            nodefactor.Domain.Health
edges                                                   0.32096407
mutual                                                  0.30198295
gwidegree                                               0.24132331
nodematch.Gender.F                                      0.21786634
nodematch.Gender.M                                      0.16632984
```

```
                                                    nodefactor.Domain.Secondary education
edges                                                                          0.50150216
mutual                                                                         0.25919743
gwidegree                                                                      0.51759758
nodematch.Gender.F                                                             0.27677143
nodematch.Gender.M                                                             0.35579405
nodematch.Gender.Org                                                           0.04130083
nodematch.Gender.Unknown                                                       0.10264946
nodefactor.Domain.Community                                                    0.05689256
nodefactor.Domain.Elementary and primary education                             0.06480525
nodefactor.Domain.Entrepreneurship                                             0.19929528
nodefactor.Domain.Government                                                   0.07293102
nodefactor.Domain.Health                                                       0.04996144
nodefactor.Domain.Higher Education                                             0.31222629
nodefactor.Domain.Languages                                                    0.22668423
nodefactor.Domain.Library                                                      0.07738162
nodefactor.Domain.Organization                                                 0.21449632
nodefactor.Domain.Other                                                        0.04483081
nodefactor.Domain.Secondary education                                          1.00000000
nodefactor.Domain.Undergraduate                                                0.02832994
nodefactor.Domain.Unknown                                                      0.21770626
nodefactor.Domain.Various                                                      0.14405499
                                                    nodefactor.Domain.Undergraduate
edges                                                                0.0286332105
mutual                                                               0.0359871415
gwidegree                                                            0.0248898312
nodematch.Gender.F                                                   0.0072848314
nodematch.Gender.M                                                   0.0490683749
nodematch.Gender.Org                                                -0.0265304955
nodematch.Gender.Unknown                                             0.0142673342
nodefactor.Domain.Community                                          0.0215442219
nodefactor.Domain.Elementary and primary education                   0.0839763396
nodefactor.Domain.Entrepreneurship                                   0.0174945243
nodefactor.Domain.Government                                        -0.0003852756
nodefactor.Domain.Health                                            -0.0184393271
nodefactor.Domain.Higher Education                                  -0.0038743172
nodefactor.Domain.Languages                                          0.0432110346
nodefactor.Domain.Library                                           -0.0032870108
nodefactor.Domain.Organization                                      -0.0146574750
nodefactor.Domain.Other                                              0.0264891384
nodefactor.Domain.Secondary education                                0.0283299379
nodefactor.Domain.Undergraduate                                      1.0000000000
nodefactor.Domain.Unknown                                            0.0200435383
nodefactor.Domain.Various                                            0.0233196842
                                                    nodefactor.Domain.Unknown
edges                                                            0.506680892
mutual                                                           0.206628708
gwidegree                                                        0.539805563
nodematch.Gender.F                                               0.249892747
nodematch.Gender.M                                               0.158205920
nodematch.Gender.Org                                             0.024801558
nodematch.Gender.Unknown                                         0.721435915
nodefactor.Domain.Community                                      0.027932339
nodefactor.Domain.Elementary and primary education               0.048084921
nodefactor.Domain.Entrepreneurship                               0.181561150
nodefactor.Domain.Government                                     0.130308433
nodefactor.Domain.Health                                         0.200517365
nodefactor.Domain.Higher Education                               0.266326684
nodefactor.Domain.Languages                                      0.227681230
nodefactor.Domain.Library                                        0.050136822
nodefactor.Domain.Organization                                   0.092549319
nodefactor.Domain.Other                                         -0.006395582
nodefactor.Domain.Secondary education                            0.217706263
nodefactor.Domain.Undergraduate                                  0.020043538
nodefactor.Domain.Unknown                                        1.000000000
nodefactor.Domain.Various                                        0.092764740
```

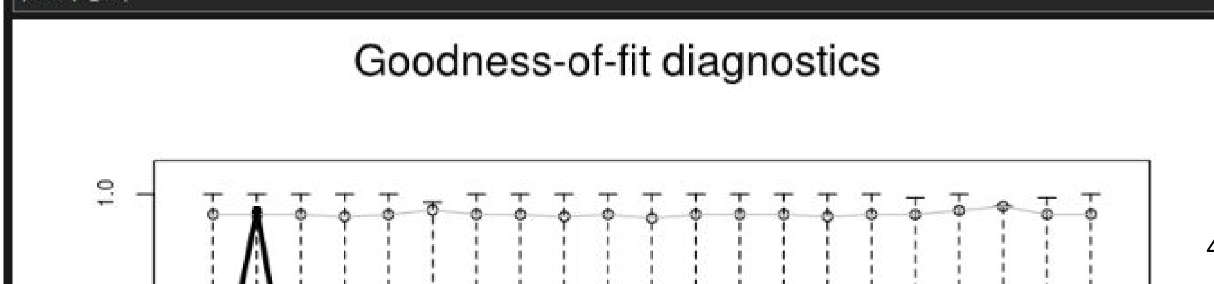
```
         nodefactor.Domain.Organization nodefactor.Domain.Other
Lag 0                         1.0000000               1.0000000
Lag 1024                      0.9255489               0.7845069
Lag 2048                      0.8606799               0.6127741
Lag 3072                      0.8023264               0.4634444
Lag 4096                      0.7565444               0.3577600
Lag 5120                      0.7191442               0.2861780
         nodefactor.Domain.Secondary education nodefactor.Domain.Undergraduate
Lag 0                                1.0000000                       1.0000000
Lag 1024                             0.8939031                       0.6684404
Lag 2048                             0.8035353                       0.4314014
Lag 3072                             0.7293580                       0.2795459
Lag 4096                             0.6684204                       0.1972900
Lag 5120                             0.6165003                       0.1314849
         nodefactor.Domain.Unknown nodefactor.Domain.Various
Lag 0                    1.0000000                 1.0000000
Lag 1024                 0.8716120                 0.9303217
Lag 2048                 0.7714445                 0.8724157
Lag 3072                 0.6887132                 0.8193043
Lag 4096                 0.6261163                 0.7723890
Lag 5120                 0.5727898                 0.7318261

Sample statistics burn-in diagnostic (Geweke):
Chain 1

Fraction in 1st window = 0.1
Fraction in 2nd window = 0.5

                                             edges
                                           0.05044
                                            mutual
                                           0.28372
                                          gwidegree
                                          -0.51463
                                 nodematch.Gender.F
                                           1.05342
                                 nodematch.Gender.M
                                          -0.77085
                               nodematch.Gender.Org
                                          -0.99103
                           nodematch.Gender.Unknown
                                          -0.06608
                        nodefactor.Domain.Community
                                           2.60822
nodefactor.Domain.Elementary and primary education
                                          -0.77229
                  nodefactor.Domain.Entrepreneurship
                                          -0.78821
                        nodefactor.Domain.Government
                                          -1.20365
                            nodefactor.Domain.Health
                                          -1.12079
                  nodefactor.Domain.Higher Education
                                           0.22530
                         nodefactor.Domain.Languages
                                           2.21018
                           nodefactor.Domain.Library
                                          -0.77069
                      nodefactor.Domain.Organization
                                          -0.52960
                             nodefactor.Domain.Other
                                          -0.42884
               nodefactor.Domain.Secondary education
                                          -0.71109
                     nodefactor.Domain.Undergraduate
                                           2.38083
                           nodefactor.Domain.Unknown
                                          -0.46930
                           nodefactor.Domain.Various
                                          -2.66171

Individual P-values (lower = worse):
                                             edges
                                       0.959769054
                                            mutual
                                       0.776626526
```

***sample stats show in first appendix, removed for brevity ***

```
# generate compare plot for goodness of fit
mgof <- gof(model_v3, GOF=~model)
plot(mgof)
```




```
# look at degree
mgof.degree <- gof(model_v3, GOF=~idegree)
plot(mgof)
```



```
# check the odds of the data
lapply(model_v3[1], exp)
```

$coef =

edges
    0.00391661826485624
mutual
    41.0547372320082
gwidegree
    0.063819727764621
nodematch.Gender.F
    1.18799674927322
nodematch.Gender.M
    1.21875251802766
nodematch.Gender.Org
    0.770151522368293
nodematch.Gender.Unknown
    4.89437985237417
nodefactor.Domain.Community
    3.51890637065117
nodefactor.Domain.Elementary and primary education
    0.868945447973084
nodefactor.Domain.Entrepreneurship
    1.40750648049384
nodefactor.Domain.Government
    0.766535273380976
nodefactor.Domain.Health
    1.45876545194954
nodefactor.Domain.Higher Education
    1.26808114702961
nodefactor.Domain.Languages
    1.4658360717404
nodefactor.Domain.Library
    0.896620272033604
nodefactor.Domain.Organization
    1.38271530297064
nodefactor.Domain.Other
    0.720486811932691
nodefactor.Domain.Secondary education
    0.985729682828447
nodefactor.Domain.Undergraduate
    0.160901322073474
nodefactor.Domain.Unknown
    0.670788002542859
nodefactor.Domain.Various
    1.25107529122044

References

CCK11 Dataset part of Gephi https://gephi.org/users/download/

Douglas A Luke, A User's Guide to Network Analysis in R

J. K. Harris, An Introduction to Exponential Random Graph Modeling

ERGM tutorial http://www.irrodl.org/index.php/irrodl/article/view/2170/3388