

# Case\_Studies\_In\_Industrial\_Modelling\_Project

Jayaram Gokulan

2/27/2020

## DESCRIPTIVES

### A. Density of Network

```
## density
## 0.01650837
```

COMMENT:

The directed network, gangs, has a density of 0.01650837. This density is towards a lower value(<0.5) indicating the sparse nature of the network.

This is alternatively calculated using:

A. The total no of edges in the network given below :

```
## edges
## 73
```

B. The total no of vertices in the network given below:

```
## [1] 67
```

The density is then = (Total No of Edges) / (Total No of Vertices)\*(Total No of Vertices-1)

$$= 73 / (67 \times 66)$$

$$= 0.01650837$$

### B. Isolates of Network

The isolates of this network are as follows :

```
## [1] 5 6 10 11 20 21 25 26 32 34 35 45 46 53 54 59 65 66
```

The number of isolates in this network are :

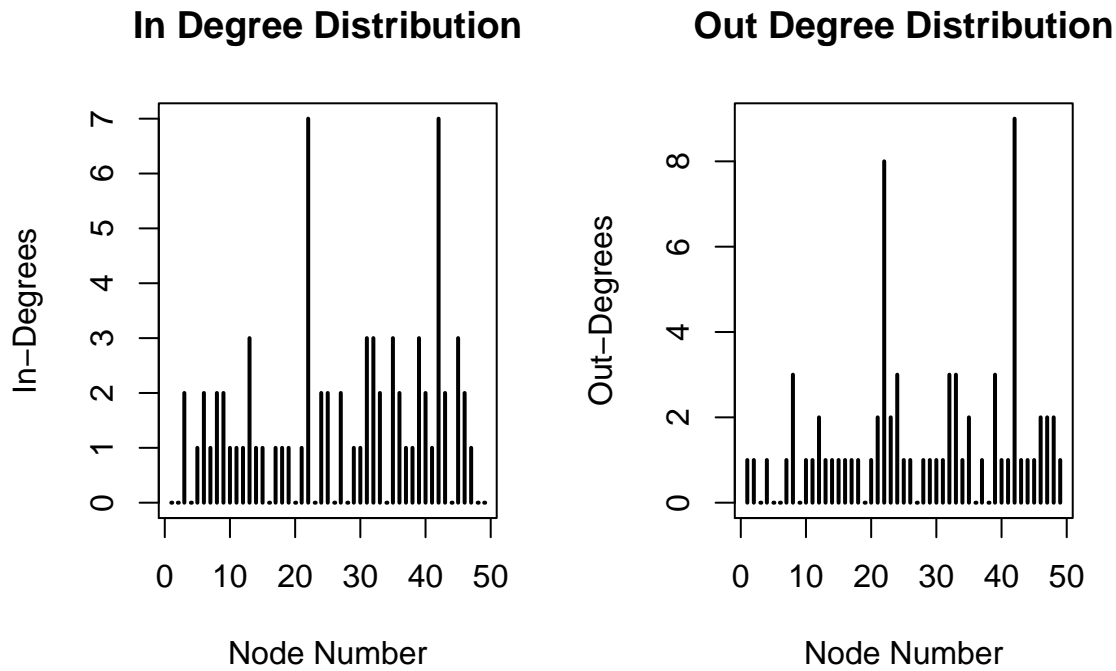
```
## [1] 18
```

COMMENT:

Thus without considering the isolates, this network has  $67 - 18 = 49$  non isolate nodes.

## B. In & Out Degree Distribution

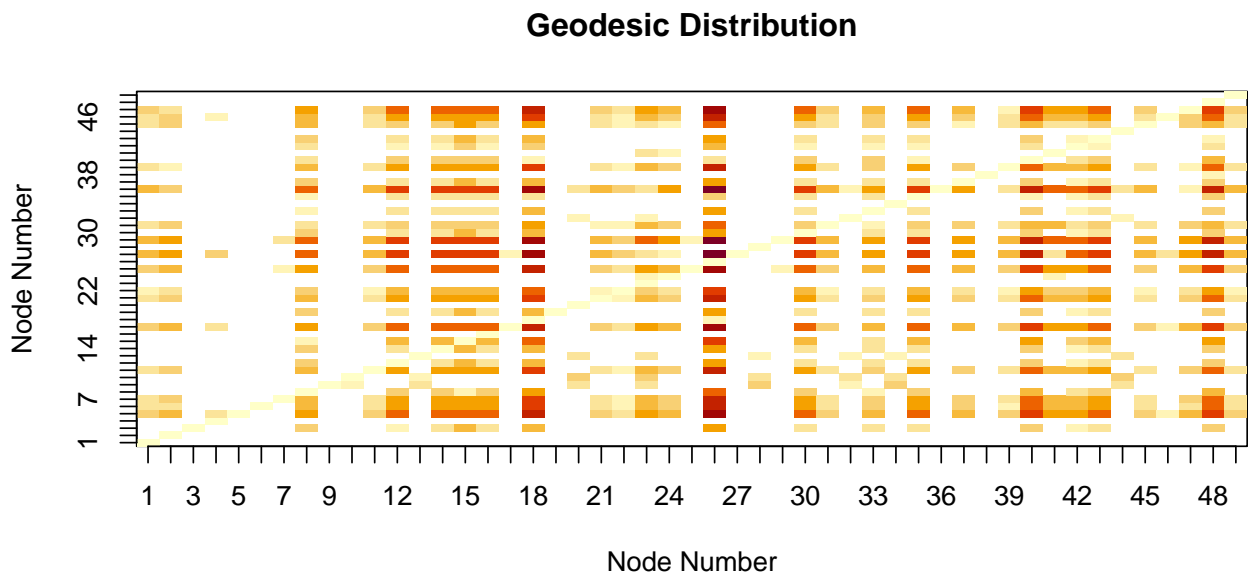
The In & Out Degree Histogram Distribution is as below :



COMMENT:

From the histogram it can be inferred that criminals with nodes 22 & 42 have the highest In Degrees of 7 each and the same pair of criminals also have the highest Out Degrees of 8 & 9. Also note that we have removed isolates when plotting this graph, as isolates do not have an in degree or outdegree.

## C. Geodesic Distribution



COMMENT :

A geodesic image shows a relative colour patten depending on the no of egdes between the shortest path connecting a pair of nodes. A darker shade usually indicates a larger no of nodes between the nodes while a lighter shade indicates the opposite.

INTERPRETATION:

The geodesic image shows that the shortest path between Nodes 29 & 26 has the highest no of edges, similarly for those between 26 & 36. Intuitively, the shortest path having the least no of edges would be from each node to itself, thus has geodesic distance of zero. Additionally the largest geodesic distance is found to be 5.

The maximum geodesic distance is alternatively calculated with the following R code:

```
lc<- component.largest(yd, result="graph")
gd<- geodist(lc)
max(gd$gdist)
```

```
## [1] 5
```

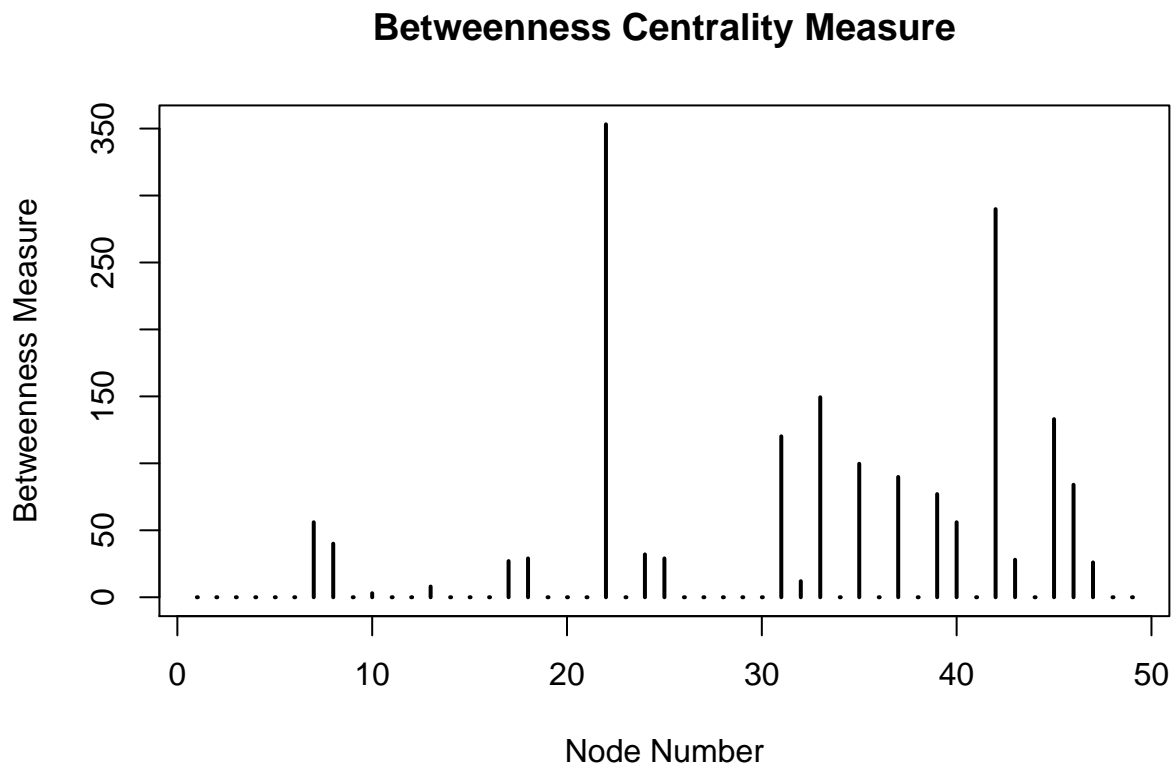
The minimum geodesic distance is alternatively calculated with the following R code :

```
lc<- component.largest(yd, result="graph")
gd<- geodist(lc)
min(gd$gdist)
```

```
## [1] 0
```

## D. Betweenness

The Betweenness Centrality Measure histogram distribution of the network is as below:



The corresponding Betweenness centrality matrix is as follows :

```
## [1] 0.00000 0.00000 0.00000 0.00000 0.00000 0.00000 0.00000 56.00000
## [8] 40.00000 0.00000 3.00000 0.00000 0.00000 0.00000 8.00000 0.00000
## [15] 0.00000 0.00000 27.00000 29.00000 0.00000 0.00000 0.00000 0.00000
## [22] 353.16667 0.00000 32.00000 29.00000 0.00000 0.00000 0.00000 0.00000
## [29] 0.00000 0.00000 120.16667 12.00000 149.33333 0.00000 99.66667
## [36] 0.00000 89.83333 0.00000 77.00000 56.00000 0.00000 289.83333
## [43] 28.00000 0.00000 133.00000 84.00000 26.00000 0.00000 0.00000
```

COMMENT :

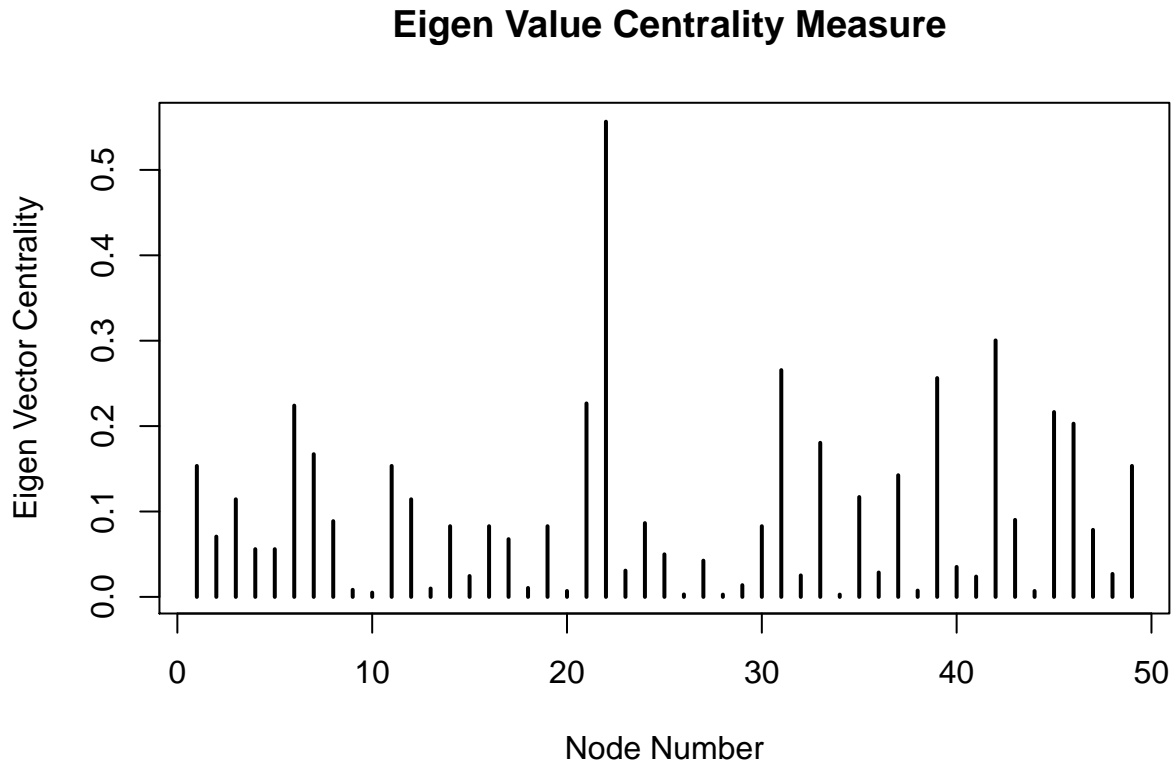
The betweenness histogram shows that the node 22 has the highest betweenness measure of 353.16667 followed by node 42 with betweenness measure of 289.83333, while the node 10 have the lowest non zero measure of 3.

INTERPRETATION:

Nodes 22 and 42, with high betweenness measure, act as ‘bridges’ in this network. They have the highest no of geodesics passing through them and hence are in a position to influence the flow around the system. The opposite can be said about node 10.

## E. Eigen Values

The Eigen Value Centrality Measure histogram distribution of the network is as below:



The corresponding Eigen Value Centrality matrix is as follows :

```
## [1] 0.153376840 0.070621210 0.114289441 0.055905635 0.055905635 0.223998049
```

```
## [7] 0.167116755 0.088694489 0.008314042 0.004993754 0.153376840 0.114289441
## [13] 0.009803631 0.082787967 0.024446765 0.082787967 0.067605775 0.010473330
## [19] 0.082787967 0.006937619 0.226572460 0.556461377 0.030752400 0.086401611
## [25] 0.049849326 0.002886753 0.042448887 0.002702164 0.013739915 0.082787967
## [31] 0.265558579 0.025170143 0.180430352 0.002702164 0.116912390 0.028573642
## [37] 0.142449146 0.007416372 0.256218446 0.035111187 0.023814781 0.300360253
## [43] 0.090204339 0.006937619 0.216454723 0.202829363 0.078496931 0.026907090
## [49] 0.153376840
```

COMMENT:

Both the graph and matrix shows that node 22 has the highest Eigen Value Centrality Measure of 0.55646 while the nodes 28 & 34 have the lowest (non zero) Eigen Value Centrality Measure of 0.002702164.

INTERPRETATION:

Thus Node 22 can be considered to be have the highest influence within this network i.e. have many/or important nodes connected to it. The opposite can be said about nodes 28 & 34.

## F. Clustering Coefficient

The clustering coefficient of the network is given as follows :

```
## [1] 0.04294479
```

COMMENT & INTERPRETATION:

Since the value of clustering coefficients always range between 0 & 1, one can infer that a value of 0.0429 as obtained above is a lower value, hence there is not much of clustering of nodes observed in this network.

## G. Reciprocity

The reciprocity measure of this a network for criminal nodes having the same nationality is given as below:

```
## Warning: `set_attr()` is deprecated as of rlang 0.3.0
## This warning is displayed once per session.
```

```
## [1] 0.1917808
```

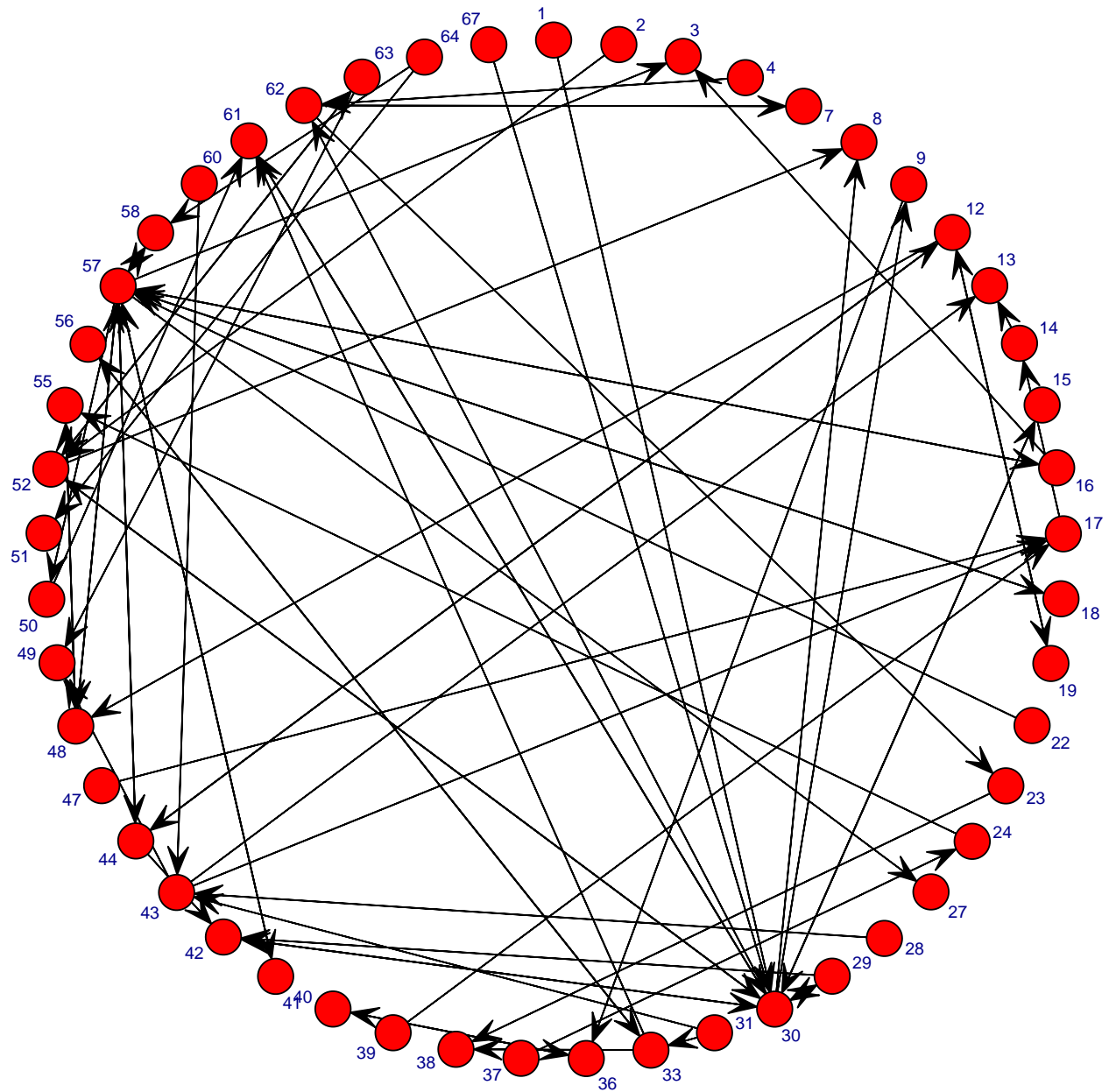
COMMENT & INTERPRETATION :

The value denotes that near to 19.17% of all nodes with same nationality have mutual ties between them.

# VISUALIZATION

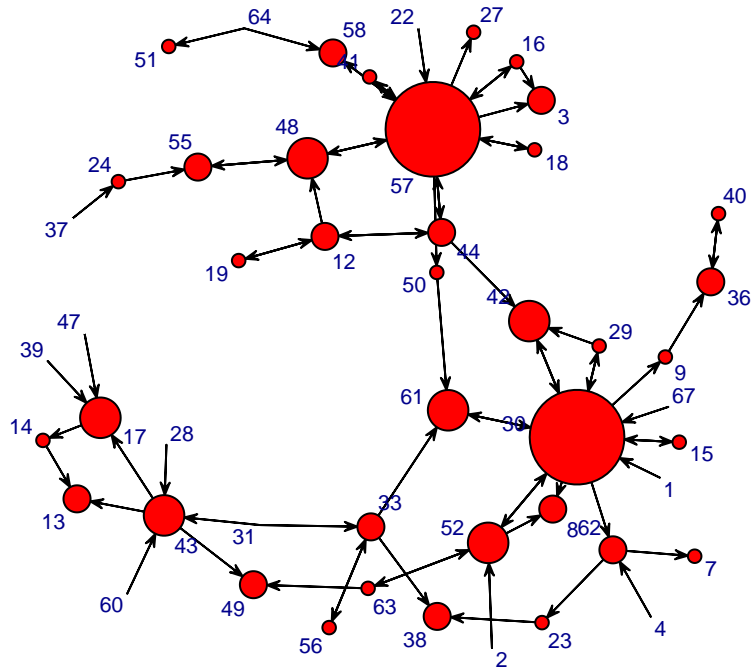
## A. Network graph using the the “circle” layout

**Gang Network in Circle Layout**

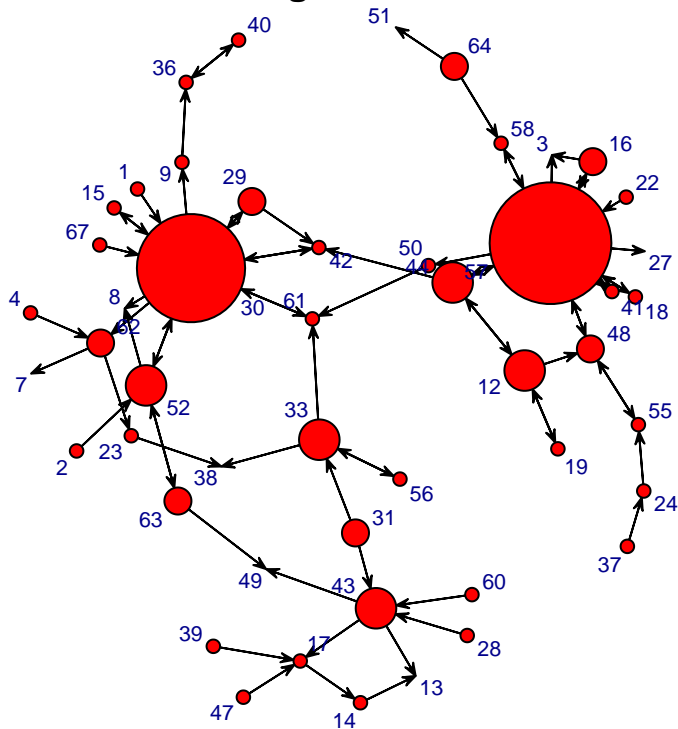


B. Network graph using the “fruchtermanreingold” layout : Node size as function of In/Out Degree

**In-Degree Network**

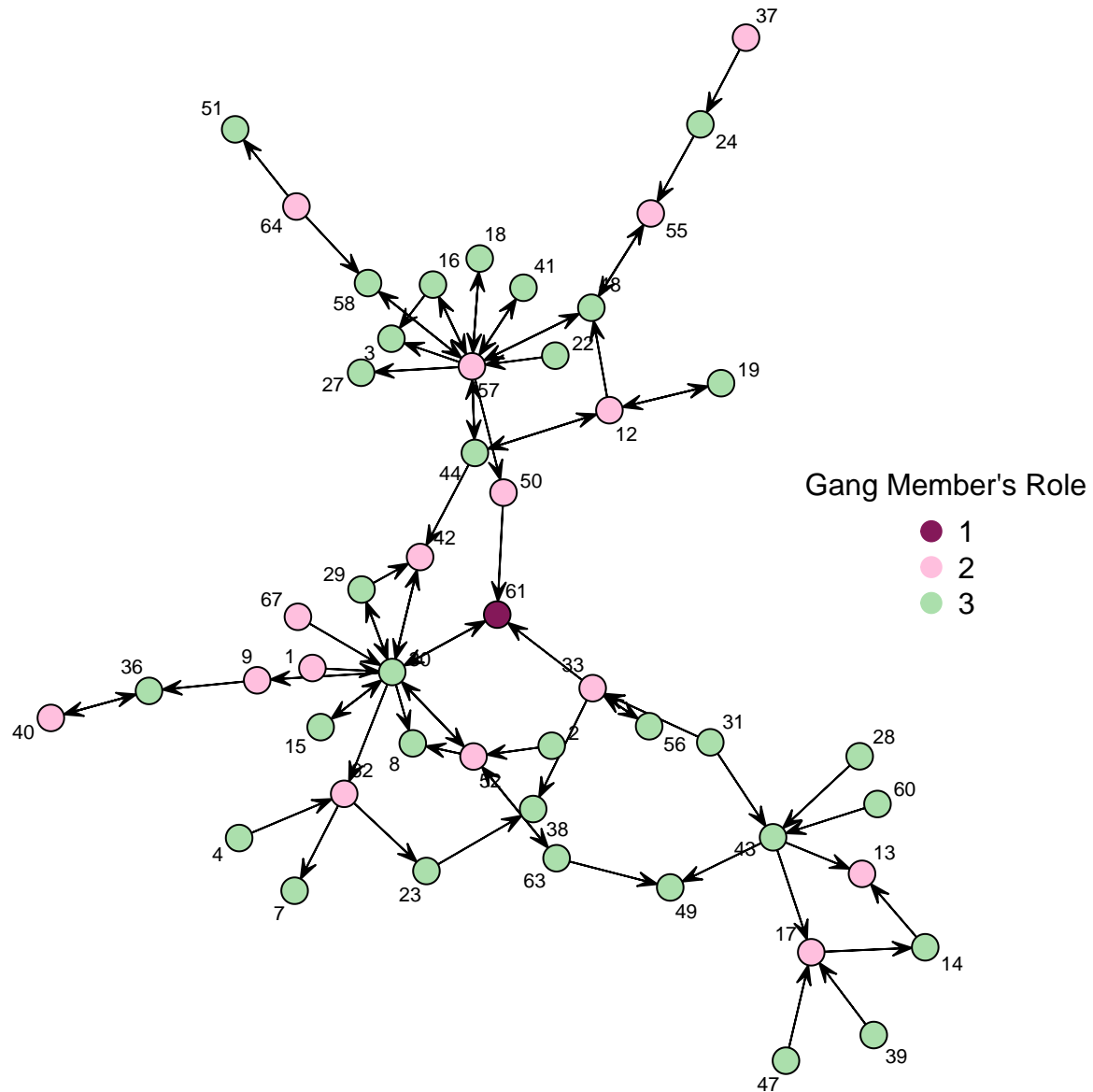


**Out-Degree Network**



C. Network graph using the “kamadakawai” layout : Node colour as function of Role

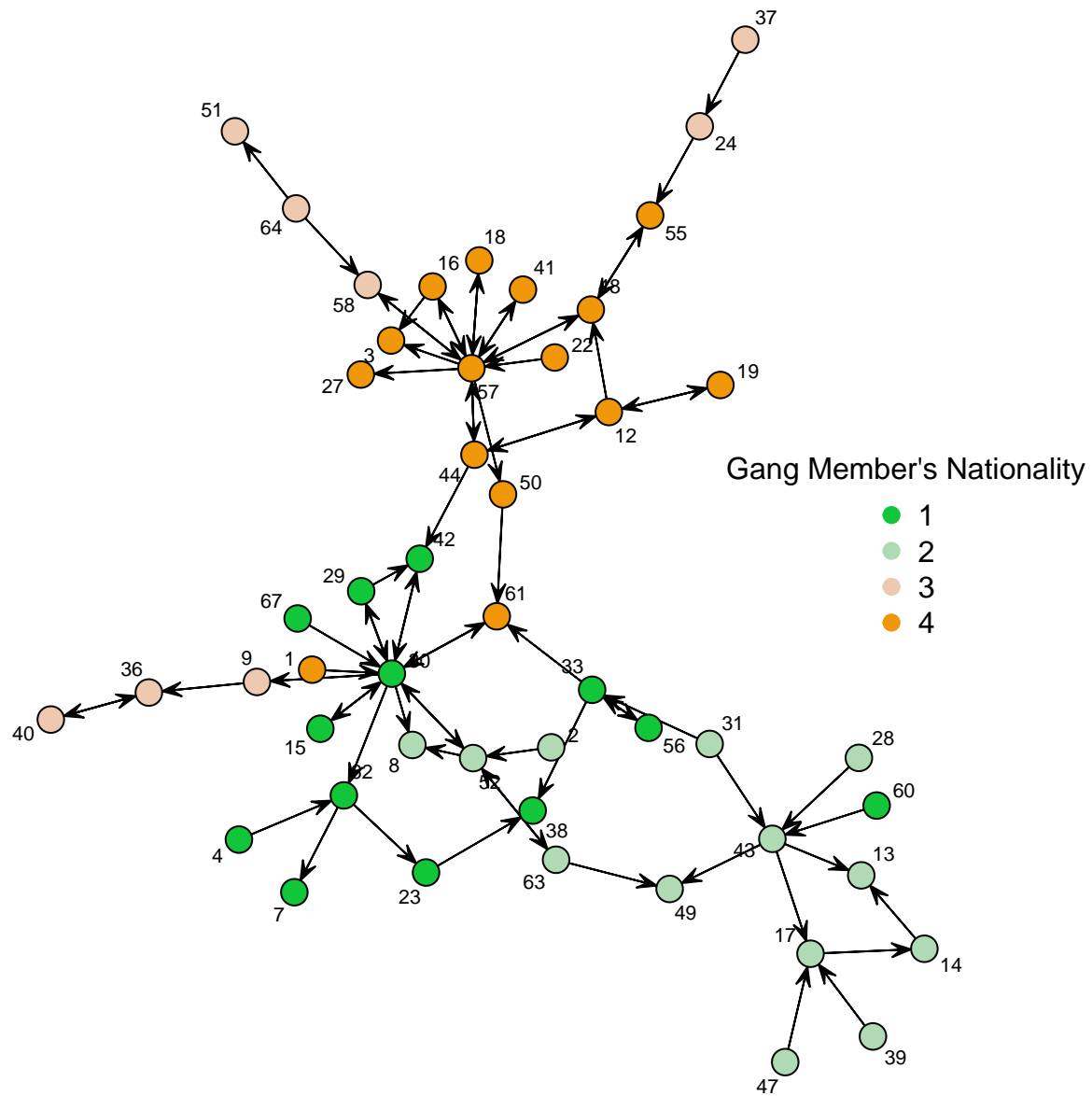
### Gang Network based on Role of Members





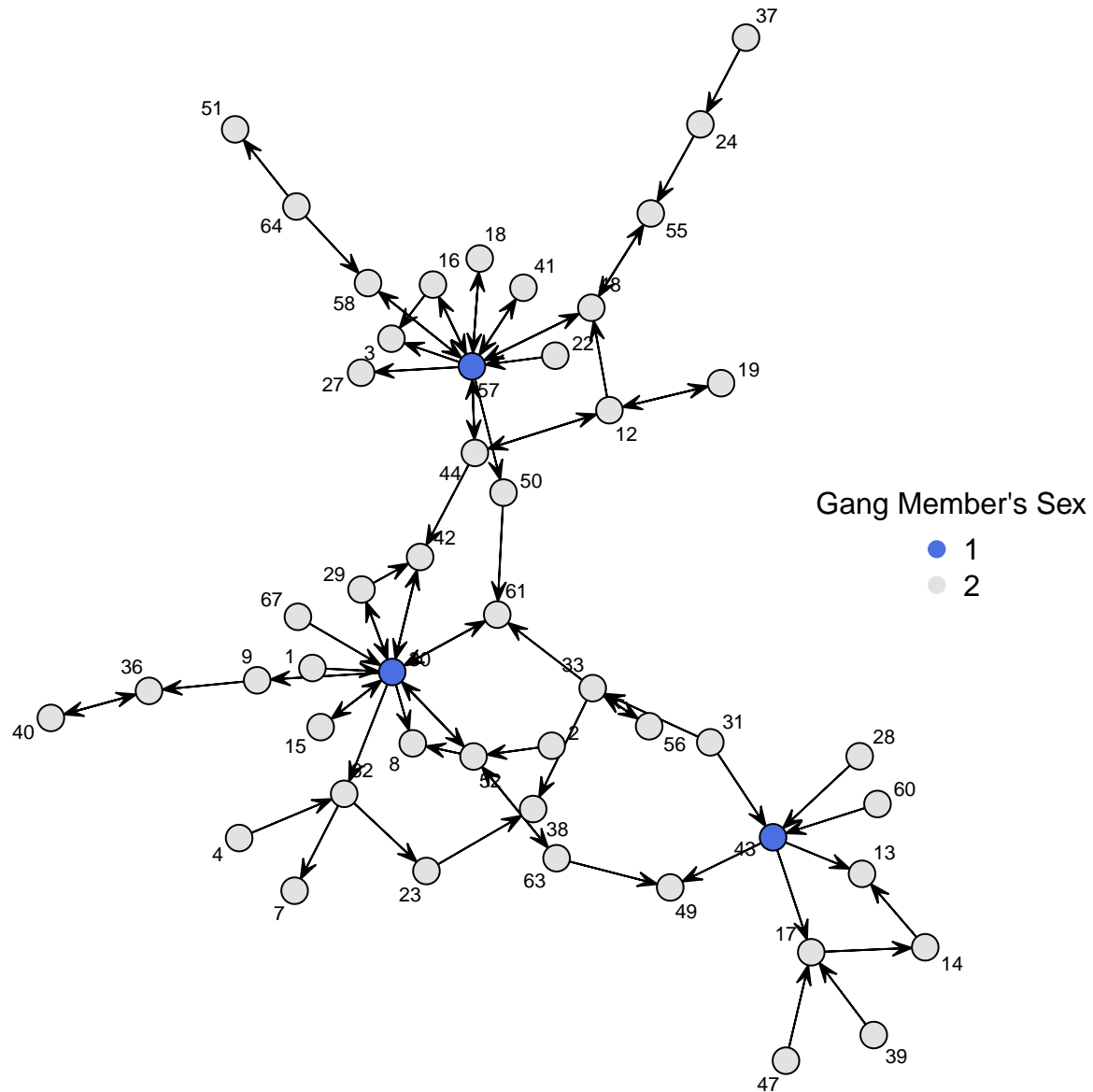
C. Network graph using the “kamadakawai” layout : Node colour as function of Nationality

### Gang Network based on Nationality of Members



C. Network graph using the “kamadakawai” layout: Node colour as function of Sex

### Gang Network based on Sex of Members



# MODELLING

## A. ERGM Model

### 1. Edge Density Only Model (Baseline Model)

```
## Starting maximum pseudolikelihood estimation (MPLE):

## Evaluating the predictor and response matrix.

## Maximizing the pseudolikelihood.

## Finished MPLE.

## Stopping at the initial estimate.

## Evaluating log-likelihood at the estimate.

##
## =====
## Summary of model fit
## =====
##
## Formula:    ynet ~ edges
##
## Iterations: 7 out of 20
##
## Monte Carlo MLE Results:
##      Estimate Std. Error MCMC % z value Pr(>|z|)
## edges   -4.087      0.118      0  -34.63  <1e-04 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##      Null Deviance: 6130  on 4422  degrees of freedom
##      Residual Deviance: 744  on 4421  degrees of freedom
##
## AIC: 746      BIC: 752.4      (Smaller is better.)
```

#### DESCRIPTION:

The log-odds of any edge occurring in this model is:

$= -4.087 \times \text{change in the number of edge}$

$= -4.087 \times 1$  (since every new edge added to the model increases the number of edge 1 )

The corresponding probability of an edge is :

$= \exp(-4.087)/(1+\exp(-4.087))$

$= 0.0165$  (which is the just the density of the network)

#### RESULT & INTERPRETATION:

- Thus the log-odds of any edge existing in the gangs network is -4.087 and reaffirms the density value we obtained from the summary statistics.
- Thus the probability of any edge existing in the gangs network is .0165

FACT:

The AIC value for this base model is 649.6.

## 2. Edge Density + Reciprocity + Reciprocity between nodes with same role + Homophily between nodes with same nationality

```
## Starting maximum pseudolikelihood estimation (MPLE):
```

```
## Evaluating the predictor and response matrix.
```

```
## Maximizing the pseudolikelihood.
```

```
## Finished MPLE.
```

```
## Starting Monte Carlo maximum likelihood estimation (MCMLE):
```

```
## Iteration 1 of at most 20:
```

```
## Optimizing with step length 1.
```

```
## The log-likelihood improved by 3.108.
```

```
## Step length converged once. Increasing MCMC sample size.
```

```
## Iteration 2 of at most 20:
```

```
## Optimizing with step length 1.
```

```
## The log-likelihood improved by 0.1718.
```

```
## Step length converged twice. Stopping.
```

```
## Finished MCMLE.
```

```
## Evaluating log-likelihood at the estimate. Using 20 bridges: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
```

```
## This model was fit using MCMC. To examine model diagnostics and check
```

```
## for degeneracy, use the mcmc.diagnostics() function.
```

```
##
## =====
## Summary of model fit
## =====
##
## Formula:   ynet ~ edges + mutual + mutual(same = "role") + nodematch("nationality")
##
## Iterations: 2 out of 20
##
## Monte Carlo MLE Results:
##
##           Estimate Std. Error MCMC % z value Pr(>|z|)
## edges           -5.6135     0.2554      0 -21.982 < 1e-04 ***
## mutual            4.6335     0.4373      0  10.595 < 1e-04 ***
## mutual.role       -2.4263     0.6678      0   -3.633 0.00028 ***
## nodematch.nationality 1.9752     0.2860      0   6.906 < 1e-04 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##      Null Deviance: 6130.2 on 4422 degrees of freedom
## Residual Deviance: 555.8 on 4418 degrees of freedom
##
## AIC: 563.8    BIC: 589.4    (Smaller is better.)
```

#### DESCRIPTION :

Conditional log-odds of two nodes forming a edge is:

$-5.6135 \times (\text{change in the number of edges}) + 4.6335 \times (\text{change in number of mutual edges}) + -2.4263 \times (\text{change in number of mutual edges among nodes with same role}) + 1.9752 \times (\text{change in number of edges between nodes with same nationality})$

The corresponding probability is :

$= \exp(\text{conditional log odds}) / (1 + \exp(\text{conditional log odds}))$

#### RESULT :

- If an added edge will not create a mutual edge nor a mutual edge between nodes with same role nor an edge between nodes of same nationality, the log-odds is: -5.6135

The probability  $= \exp(-5.6135) / (1 + \exp(-5.6135)) = 0.003635016$  (which is the density of this model)

- If an added edge will create a mutual edge but not a mutual edge between nodes with same role or an edge between nodes of same nationality, the log-odds is:  $-5.6135 + 4.6335 = -0.98$

The probability is  $= \exp(-0.98) / (1 + \exp(-0.98)) = 0.2728918$

- If an added edge will only create a mutual edge between nodes with same role but not any other mutual edge or an edge between nodes of same nationality, the log-odds is:  $-5.6135 - 2.4263 = -8.0398$

The probability is  $= \exp(-8.0398) / (1 + \exp(-8.0398)) = 0.0003222695$

- If an added edge will create an edge between nodes of same nationality but not any mutual edge or a mutual edge between nodes with same role, the log-odds is:  $-5.6135 - 1.9752 = -7.5887$

The probability is  $= \exp(-7.5887) / (1 + \exp(-7.5887)) = 0.0005058826$

- If an added edge will create a mutual edge, specifically a mutual edge between nodes with same role but not an edge between nodes of same nationality, the log-odds is:  $-5.6135 + 4.6335 - 2.4263 = -3.4063$

The probability is  $= \exp(-3.4063)/(1+\exp(-3.4063)) = 0.03209915$

- If an added edge will create a mutual edge and an edge between nodes of same nationality but not a mutual edge between nodes with same role , the log-odds is:  $-5.6135 + 4.6335 - 1.9752 = -2.9552$

The probability is  $= \exp(-2.9552)/(1+\exp(-2.9552)) = 0.06918992$

- If an added edge will create a mutual edge between nodes with same role and an edge between nodes of same nationality but not any other mutual edge, the log-odds is:  $-5.6135 - 2.4263 - 1.9752 = -10.015$

The probability is  $= \exp(-10.015)/(1+\exp(-10.015)) = 4.472201e-05$

- If an added edge will simultaneously create a mutual edge, mutual edge between nodes with same role and an edge between nodes of same nationality, the log-odds is:  $-5.6135 + 4.6335 - 2.4263 - 1.9752 = -5.3815$

The probability is  $= \exp(-10.015)/(1+\exp(-10.015)) = 0.004579844$

INTERPRETATION :

- The density of this model is given by 0.003635016.
- Much of the density is captured by mutual edges hence reflecting higher probability (0.272898) for an added edge to have reciprocity.
- Some of the density is captured by mutual edges which are between nodes of same nationality hence reflecting slightly higher probability of 0.06918992.
- Some of the density is captured by mutual edges between nodes of same role hence reflecting slightly higher probability of 0.03209915.
- Based on the probability values obtained, a significantly lower portion of the density accounts for an added edge that will be a mutual edge between nodes of same role and edge between nodes of same nationality.
- Based on the probability values obtained, a significantly lower portion of the density accounts for an added edge that will be a mutual edge between nodes of same role and edge between nodes of same nationality.
- Based on the probability values obtained, a significantly lower portion of the density accounts for an added edge that will be a mutual edge between nodes of same role.
- Based on the probability values obtained, a significantly lower portion of the density accounts for an added edge that will be an edge between nodes of same nationality.

FACT: Also the AIC value for this base model is 563.8 indicating that this is a better model compared to the baseline model.

## B. Goodness of Fit

```
##
## Goodness-of-fit for in-degree
##
##   obs min  mean max MC p-value
## 0   30  12 22.41  34    0.12
## 1   17  14 24.58  35    0.08
## 2   12   5 13.32  24    0.86
## 3    6   1  4.74  12    0.72
## 4    0   0  1.60   8    0.48
## 5    0   0  0.25   3    1.00
```

```

## 6  0  0  0.06  1      1.00
## 7  2  0  0.04  1      0.00
##
## Goodness-of-fit for out-degree
##
##   obs min  mean max MC p-value
## 0  26  10 22.74  33      0.52
## 1  27  15 23.93  34      0.58
## 2   7   5 13.40  26      0.06
## 3   5   0  5.24  13      1.00
## 4   0   0  1.28   4      0.60
## 5   0   0  0.31   3      1.00
## 6   0   0  0.09   1      1.00
## 7   0   0  0.01   1      1.00
## 8   1   0  0.00   0      0.00
## 9   1   0  0.00   0      0.00
##
## Goodness-of-fit for edgewise shared partner
##
##   obs min  mean max MC p-value
## esp0  66  49 71.22  99      0.80
## esp1   7   0  2.49  13      0.16
## esp2   0   0  0.02   1      1.00
##
## Goodness-of-fit for minimum geodesic distance
##
##   obs  min   mean  max MC p-value
## 1    73   50  73.73 109      1.00
## 2   154   28  78.29 166      0.02
## 3   127   11  73.78 171      0.16
## 4   101    3  63.59 173      0.40
## 5    86    1  50.30 173      0.44
## 6    54    0  38.05 153      0.54
## 7    37    0  27.22 127      0.56
## 8    16    0  18.31 105      0.80
## 9     7    0  11.70  86      0.82
## 10    3    0   7.39  63      0.80
## 11    0    0   4.51  51      1.00
## 12    0    0   2.52  39      1.00
## 13    0    0   1.36  26      1.00
## 14    0    0   0.63  20      1.00
## 15    0    0   0.41  18      1.00
## 16    0    0   0.23  13      1.00
## 17    0    0   0.15  13      1.00
## 18    0    0   0.09   9      1.00
## 19    0    0   0.08   8      1.00
## 20    0    0   0.05   5      1.00
## 21    0    0   0.03   3      1.00
## Inf 3764 3111 3969.58 4317      0.38
##
## Goodness-of-fit for model statistics
##
##                                     obs min  mean max MC p-value
## edges                                     73  50 73.73 109      1.00

```

## mutual	17	10	17.29	29	1.00
## mutual.role	2	0	1.96	7	1.00
## nodematch.nationality	59	38	60.26	98	0.98

#### DESCRIPTION :

The simulation of the possible networks gives a sample statistical distribution. It thus necessitates us to conduct a hypothesis test to understand how confident we can be about the statistics obtained from the distribution.

The Null Hypothesis : The simulated network attribute is same as the original network attribute.

Alternate Hypothesis : The simulated network attribute is not the same as the original network attribute.

We reject the Null Hypothesis when we have evidence of a low p value i.e. when the mean of the simulated values differs a lot from that of the original network.

#### INTERPRETATION :

##### A. In-degree of nodes simulations

- It is noted that in all of the simulations, in-degree values of 5 & 6 are seen in a number of nodes much close to that found in the original observed network. This is evidenced by the near perfect high p values(=1.00), which indicates that we fail to reject the Null Hypothesis and conclude that in-degree of 5 & 6's in the original observed network are sufficiently captured in the nodes of the simulated network.
- Similarly, in-degree value of 2 and to a slightly lesser extent, that of 3 & 4 are also seen in a number of nodes (of the simulated network) close to that found in the original observed network. Again this is evidenced by slightly high p values.
- However, nodes having in-degree value of 7 in the original network are not captured in most of the simulations, since the p value for that parameter = 0.00.
- Similarly, number of simulated nodes having in-degree value of 0 & 1 are also not much similar to that found in the original network.

##### B. Out-degree of nodes simulations

- It is noted that in all of the simulations, out-degree values of 3,5,6 & 7 are seen in a number of nodes much close to that found in the original observed network. This is evidenced by the near perfect high p values(=1.00), which indicates that we fail to reject the Null Hypothesis and conclude that out-degree of 3,5,6 & 7's in the original observed network are sufficiently captured in the nodes of the simulated network.
- Similarly, out-degree value of 4 and to a slightly lesser extent, that of 0 & 1 are also seen in a number of nodes (of the simulated network) close to that found in the original observed network. Again this is evidenced by slightly high p values.
- However, nodes having out-degree value of 8 & 9 in the original network are not captured in most of the simulations, since the p value for those parameters are 0.00.
- Similarly, the number of simulated nodes having out-degree value of 2 are not much similar to that found in the original network.

##### C. Edgewise shared partner simulations

- It is noted that in all of the simulations, nodes having 2 shared partners are seen in a number of simulated nodes closest to that found in the original network. This is evidenced by the near perfect high p values(=1.00),



which indicates that we fail to reject the Null Hypothesis and conclude that number of nodes with 2 shared partner in the original network are sufficiently captured in the simulations.

- Similarly, nodes having 0 shared partners are also seen in a number of simulated nodes closer to that found in the original network. Again this is evidenced by higher p values.
- However, nodes having 1 shared partner are seen in fewer nodes than that seen in the observed network, evidenced by the low p values ( $=0.16$ ).

#### D. Minimum geodesic distance of nodes simulations

- It is noted that in all of the simulations, minimum geodesic distances ranging from 11 to 21 as well as 1 are seen in a number of nodes much close to that found in the original observed network. This is evidenced by the near perfect high p values( $=1.00$ ), which indicates that we fail to reject the Null Hypothesis and conclude that above minimum geodesic distances in the original observed network are sufficiently captured in the nodes of the simulated network.
- Similarly, minimum geodesic distance of 9,8 & 10 and to a slightly lesser extent, that of 7,6,5, 4 and even infinity are also seen across a number of nodes (of the simulated network) close to that found in the original observed network. Again this is evidenced by slightly high p values.
- However, minimum geodesic distance of 2 in the original network are not captured in most of the simulations, since the p value for that parameter is 0.02.
- Similarly, number of nodes having a minimum geodesic distance of 3 are also not much similar to that found in the original network, owing to the low p value of 0.16.

#### E. Model statistics simulations

- It is noted that in all of the simulations, number of edges is very similar to that found in the original observed network. This is evidenced by the near perfect high p values( $=1.00$ ), which indicates that we fail to reject the Null Hypothesis and conclude that edges in the original observed network are sufficiently captured in the simulated network.
- The situation is very similar for reciprocity & reciprocity between nodes having same role, where the prefect high p values( $=1.00$ ), suggest that the simulations almost always capture both these effects between the nodes.
- Like wise homophily between nodes of same nationality is also captured well by the simulated networks.