# DS745 FA19 Networking Project

## London Street Gang Network (2005-2009)

Mark Riley

11/10/2019

## Introduction to the London Street Gang Network

This network data set is a weighted network representing strengths of relationships among 54 confirmed members of a London street gang, 2005-2009. The network is undirected. Data is based on co-offenders in a London-based inner-city street gang, 2005-2009, operating from a social housing estate. Data comes from anonymised police arrest and conviction data for 'all confirmed' members of the gang. Nodes are gang members, and the edge weight gives the level of relationship:

1. Hang out together
2. Co-offend together
3. Co-offend together, serious crime
4. Co-offend together, serious crime, kin

Below we can see the number of gang member connections by relationship in the table below.

```
##
##   1   2   3   4
## 182  92  25  16
```

Most of the gang members hang out together rather than co-offend together.

Node attributes include age in years, birthplace (1. West Africa; 2. Caribbean; 3. UK; 4. East Africa), number of arrests, number of convictions, and if the gang member has served a prison sentence (0. No; 1. Yes).

For this analysis we will use the full London Gang Network data set.

## Network Analysis

### Key Characteristics

The London Gang network as a size of 54 confirmed gang members. The network density is 0.22. Density is the number of connections between gang members, divided by the maximum number of possible connections. Density values range between 0 and 1. The closer the density is to 1, the more interconnected the network.

This network has 1 component. A component is a subgroup of the network in which all gang members are connected, directly or indirectly. The network diameter is 4. The diameter is the shortest distance between the two most distant gang members in the network. The clustering coefficient is 0.52. This measures the degree to which nodes in a graph tend to cluster together.

## Gang Member Prominence

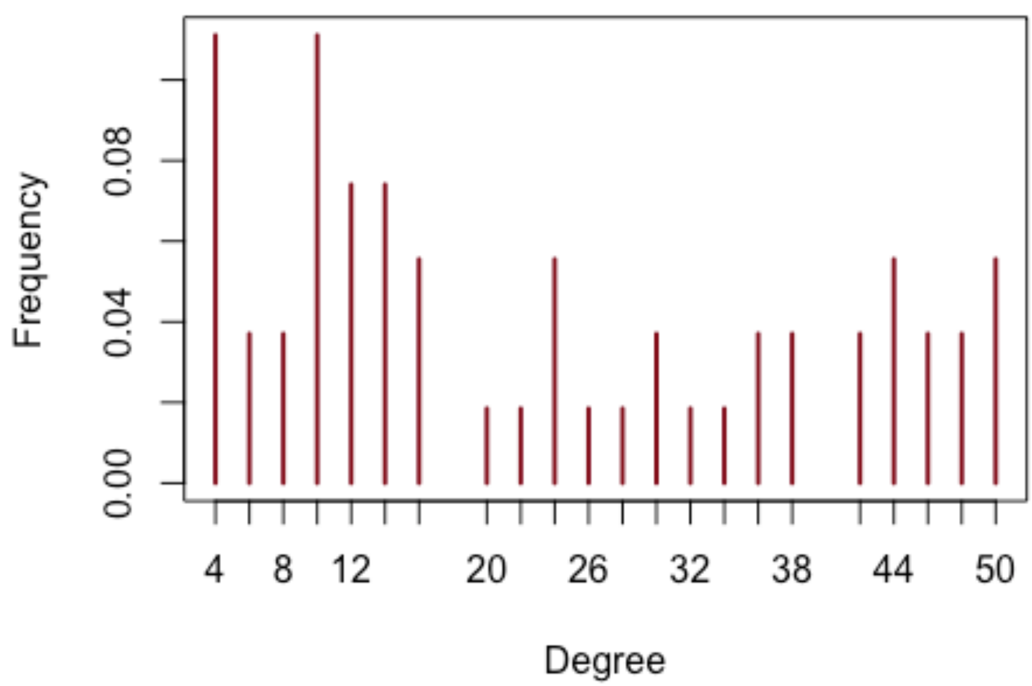
The table below shows the prominence of each member of the gang. The gang members are represented by an alias consisting of 'X' plus a number. Degree is the number of connections a gang member has to other gang members. Closeness represents the extent that a gang member is close to all other gang members in the network. Betweenness is the extent that a gang member sits between pairs of other gang members in the network, such that a path between the other members has to go through that member.

```
##                    Degree Closeness  Betweenness
## X1              50.0000000 0.6540000 299.48900000
## X7              50.0000000 0.6540000 208.06700000
## X12             50.0000000 0.6390000 164.21100000
## X14             48.0000000 0.6020000 176.45700000
## X22             48.0000000 0.6090000 124.02700000
## X23             46.0000000 0.6090000 153.63800000
## X25             46.0000000 0.6090000 153.51700000
## X2              44.0000000 0.6090000 118.07200000
## X3              44.0000000 0.6020000  87.18500000
## X10             44.0000000 0.6240000 133.24200000
## X4              42.0000000 0.6160000 199.84100000
## X9              42.0000000 0.6160000 133.00600000
## X5              38.0000000 0.6090000 132.36700000
## X21             38.0000000 0.5760000  29.60200000
## X11             36.0000000 0.5820000  95.27000000
## X28             36.0000000 0.5700000  75.82000000
## X29             34.0000000 0.5640000  38.34300000
## X6              32.0000000 0.5820000  68.84600000
## X8              30.0000000 0.5520000   6.86000000
## X20             30.0000000 0.5250000 175.50200000
## X18             28.0000000 0.5460000 122.67100000
## X19             26.0000000 0.5100000 130.36000000
## X31             24.0000000 0.4950000  12.74400000
## X35             24.0000000 0.5100000  13.70800000
## X36             24.0000000 0.5100000  13.70800000
## X13             22.0000000 0.4950000  11.48400000
## X33             20.0000000 0.4910000  10.08900000
## X16             16.0000000 0.4690000   8.09700000
## X27             16.0000000 0.5100000   0.61900000
## X34             16.0000000 0.4530000   2.01600000
## X15             14.0000000 0.4570000   2.85700000
## X32             14.0000000 0.4310000   3.94200000
## X37             14.0000000 0.4270000   3.59400000
## X51             14.0000000 0.4530000  16.25200000
## X24             12.0000000 0.4240000   1.97800000
## X26             12.0000000 0.4240000   1.97800000
## X43             12.0000000 0.4270000  26.46700000
## X54             12.0000000 0.4690000   6.64300000
## X17             10.0000000 0.4490000   0.00000000
```

```
## X30             10.0000000 0.4310000   0.18200000
## X41             10.0000000 0.4570000   1.76400000
## X42             10.0000000 0.3810000   9.51100000
## X46             10.0000000 0.4570000  13.86700000
## X48             10.0000000 0.4490000   5.82700000
## X44              8.0000000 0.4530000  20.42300000
## X52              8.0000000 0.3980000   0.00000000
## X47              6.0000000 0.4210000   1.08800000
## X49              6.0000000 0.4050000   0.00000000
## X38              4.0000000 0.4110000   0.00000000
## X39              4.0000000 0.3930000   0.00000000
## X40              4.0000000 0.3560000   0.00000000
## X45              4.0000000 0.4020000   0.00000000
## X50              4.0000000 0.3560000   0.00000000
## X53              4.0000000 0.3680000   0.77200000
## Centralization  0.2612482 0.1584381   0.09007052
```

From the table above, we can see that a high number of connections for a gang member correlates to high levels of closeness and betweenness. Gang members X1, X7, and X12 are the three most prominent members of the London Street Gang.

## Visualizations




*Degree Frequency of London Street Gang Network*

The figure above plots the frequency of the degrees in the London Street Gang Network data. Members with four and ten degrees are the most frequent. No gang member has fewer than four degrees and no more than 50.


*London Street Gang Network Diagram*

The London Street Gang network is picture above. The gang members are labeled with their aliases (e.g. X1). The color of each member corresponds with his place of birth. The thickness of the connection between the nodes indicates the type of relationship between the gang members, with the thinnest connections corresponding to hanging out and the thickest corresponding to co-offenders, serious crime, kin.

The graph depicts the members with the highest prominence in the center of the network (e.g. X1, X7, X12). The members with the lowest prominence (e.g. X45, X50, X53) at the periphery of the network. The majority of the connections are thin, corresponding to a hanging out relationship rather than co-offending.

## Community Detection

Community detection in networks is the identification of subgroups among the gang members. The subgroups are characterized by a large number of internal connections between the subgroup members, and also relatively few ties from the subgroup to other parts of the gang.

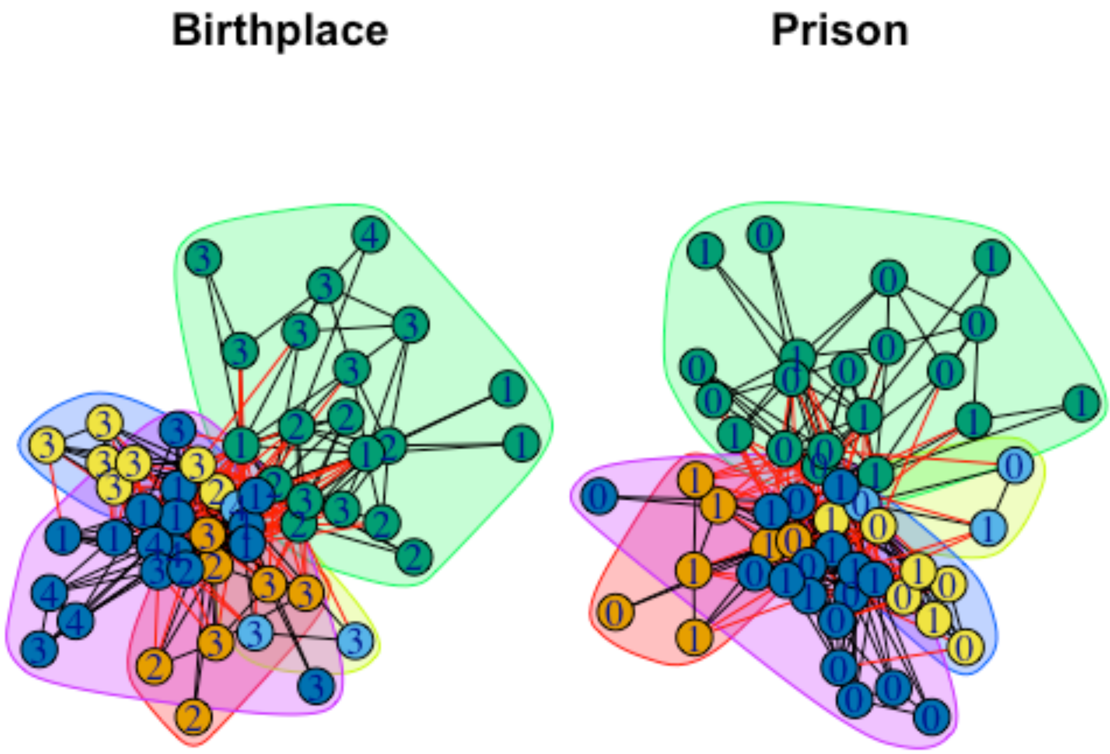
We used a number of community detection algorithms to identify subgroups in the London Street Gang. The characteristics of the network data set informed our decision on which algorithms to select. The London Street Gang network is undirected, weighted, and has a single component. These characteristics fit the Walktrap, Edge-Betweenness, Spinglass, Fast-Greedy, and Louvain algorithms.

```
##                Modularity
## Birthplace    0.082242378
## Age          -0.003613001
## Arrests      -0.020372890
## Convictions  -0.008339632
## Prison        0.007432603
## Relationship -0.023507181
```

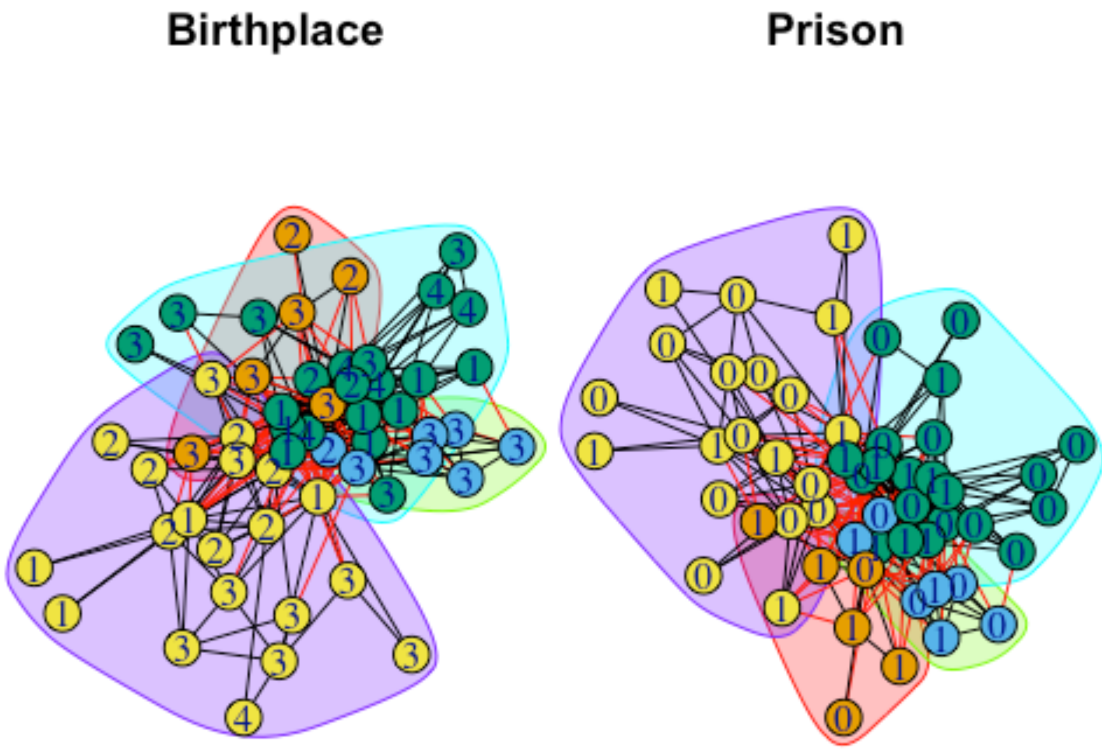
The table above shows the modularity scores for each of the node and edge attributes of the network. A positive modularity value suggests that the attribute explains some of the clustering present in the network. We see that Birthplace and Prison are the only two attributes with positive modularity.

```
##                  Modularity
## Walktrap         0.16690350
## Edge-Betweenness 0.05272361
## Spinglass        0.26147644
## Fast-Greedy      0.25594356
## Louvain          0.24856639
```

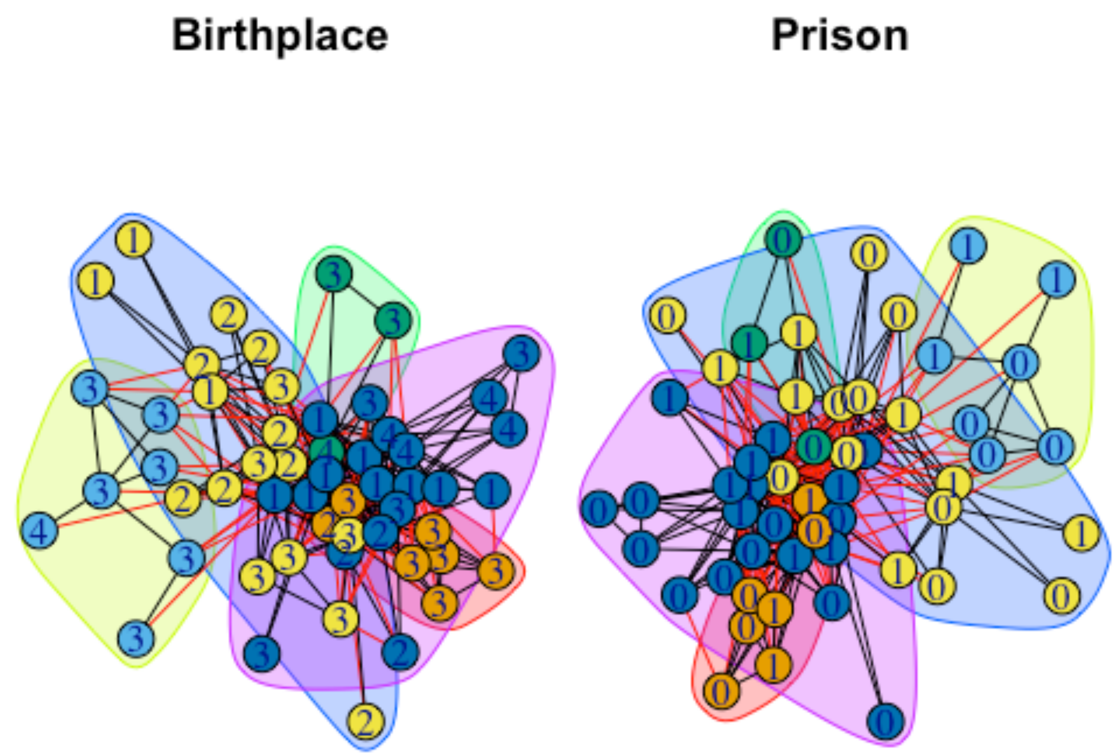
The table above displays the results of the community detection algorithms in the form of a modularity score. The top three performing algorithms were Spinglass (six subgroups), Fast-Greedy (four subgroups), and Louvain (five subgroups). We will plot each of the top three results below. The values in each circle will represent either the Birthplace or Prison values for that gang member, since those were the two attributes with positive modularities.




*Spinglass Community Detection*




*Fast-Greedy Community Detection*




*Louvain Community Detection*

The Spinglass algorithm performed slightly better than Fast-Greedy and Louvain. That said, we like that the Louvain (five subgroups) has three subgroups where members come from only two Birthplace locations. The fourth subgroup has members from three Birthplace locations and the remaining subgroup has members from all of the Birthplace locations.

```
##
##     1 2 3 4 5
##   1 0 0 0 3 9
##   2 1 0 0 8 3
##   3 6 6 2 6 4
##   4 0 1 1 0 4
```

When looking at the Prison values across the Louvain subgroups, each subgroup has roughly even numbers of members who have, and have not, been to Prison. This corresponds well with the edge attributes showing roughly half of the relationships between members represent hanging out vs. co-offending.

```
##
##      1  2  3  4  5
##   0  4  4  2 10 10
##   1  3  3  1  7 10
```

## Network Modeling

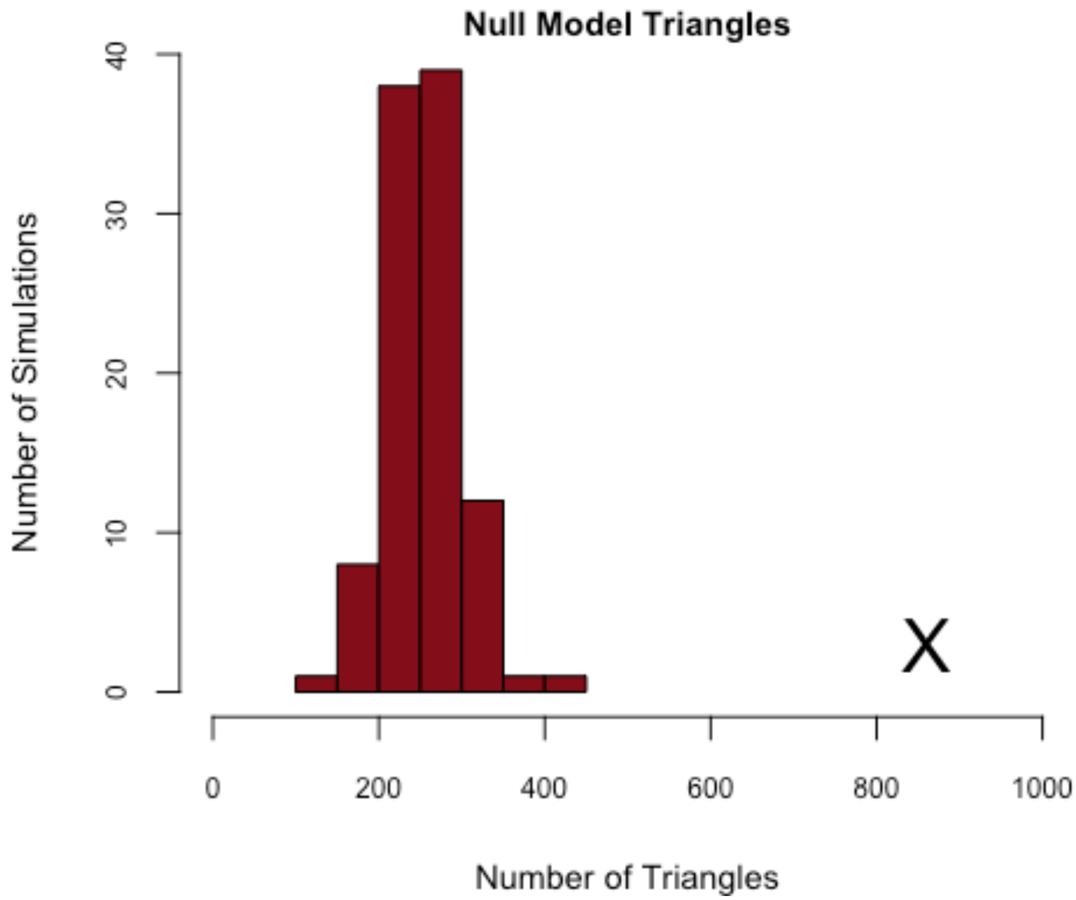
We will now perform some network modeling using exponential random graph models (ERGMs). ERGMs are a flexible and powerful approach to build and test statistical models of networks.

### NULL Model

We will begin with the NULL model, which is the simplest model using only the network edges.

```
##
## ==========================
## Summary of model fit
## ==========================
##
## Formula:   london ~ edges
##
## Iterations:  5 out of 20
##
## Monte Carlo MLE Results:
##       Estimate Std. Error MCMC % z value Pr(>|z|)
## edges  -1.2649     0.0638      0  -19.83   <1e-04 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##      Null Deviance: 1984  on 1431  degrees of freedom
##  Residual Deviance: 1508  on 1430  degrees of freedom
##
## AIC: 1510    BIC: 1516    (Smaller is better.)
```

The NULL Model coefficient of edges is negative (-1.265). This shows that the density of the network is less than 50%, which lines up with the calculated density (0.22) from the network description.




The histogram above shows that 100 simulated networks based on the NULL Model are not able to capture how triangles are formed in the network. There are 860 triangles in the London Street Gang network, but the simulations based on the NULL Model we developed show very few triangles (mean of 254.87).

## Adding Node Attributes

Based on the community detection performed above, we know that two node attributes have positive modularities - meaning they account for some of of the clustering present in the network. Those attributes were Birthplace and Prison, in decreasing order. It might be reasonable to assume that gang members are more likely to be connected based on those attributes over the NULL model.

```
##
## ==========================
## Summary of model fit
## ==========================
##
## Formula:   london ~ edges + nodefactor("Birthplace") +
nodefactor("Prison")
##
## Iterations:  5 out of 20
##
```

```
## Monte Carlo MLE Results:
##                          Estimate Std. Error MCMC % z value Pr(>|z|)
## edges                    -0.70761    0.18592      0  -3.806 0.000141 ***
## nodefactor.Birthplace.2 -0.30775    0.13177      0  -2.336 0.019512 *
## nodefactor.Birthplace.3 -0.70469    0.11920      0  -5.912  < 1e-04 ***
## nodefactor.Birthplace.4 -0.11524    0.15773      0  -0.731 0.465013
## nodefactor.Prison.1      0.20035    0.09321      0   2.149 0.031596 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##      Null Deviance: 1984  on 1431  degrees of freedom
##  Residual Deviance: 1462  on 1426  degrees of freedom
##
## AIC: 1472    BIC: 1499    (Smaller is better.)
```

The results show that the Birthplace factors, except for 4-West Africa, and factor of Prison are significantly associated with the likelihood of observing a connection between two gang members.

```
##                        edges triangle
## London Gangs             315      860
## NULL Model               292      195
## Birthplace and Prison    297      260
```

Simulation analysis based on 100 simulations each for NULL model and adding node attributes of Birthplace and Prison does improve upon the NULL model. The number of edges and triangles are slightly closer to the actual network. However the best simulated model is still very far from capturing the true number of triangles.

```
##                              AIC
## NULL Model              1510.466
## Birthplace and Prison 1472.498
```

Analysis of the AIC values shows that the model using the node attributes has better performance than the NULL model.

## Adding Homophily Effect

We will now test if connections between gang members are more or less likely based on shared Birthplace.

```
##
## ==========================
## Summary of model fit
## ==========================
##
## Formula:   london ~ edges + nodematch("Birthplace") + nodefactor("Prison")
##
## Iterations:  5 out of 20
##
## Monte Carlo MLE Results:
```

```
##                       Estimate Std. Error MCMC % z value Pr(>|z|)
## edges                 -1.57920    0.11844      0 -13.333  < 1e-04 ***
## nodematch.Birthplace  0.36545    0.13568      0   2.693  0.00707 **
## nodefactor.Prison.1   0.21658    0.09198      0   2.355  0.01854 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##      Null Deviance: 1984  on 1431  degrees of freedom
##  Residual Deviance: 1496  on 1428  degrees of freedom
##
## AIC: 1502    BIC: 1518    (Smaller is better.)
```

The model above uses the Birthplace attribute of each gang member to assess the affect on the likelihood of a connection when both gang members were born in the same location. The model shows that the Birthplace match parameter is positive and statistically significant. This indicates there is a homophily effect here.

```
##
## ==========================
## Summary of model fit
## ==========================
##
## Formula:   london ~ edges + nodematch("Birthplace", diff = TRUE) +
nodefactor("Prison")
##
## Iterations:  5 out of 20
##
## Monte Carlo MLE Results:
##                         Estimate Std. Error MCMC % z value Pr(>|z|)
## edges                   -1.57425    0.11967      0 -13.155  < 1e-04 ***
## nodematch.Birthplace.1  1.52006    0.25976      0   5.852  < 1e-04 ***
## nodematch.Birthplace.2  0.73405    0.27069      0   2.712  0.00669 **
## nodematch.Birthplace.3 -0.21831    0.17997      0  -1.213  0.22511
## nodematch.Birthplace.4  1.49695    0.52464      0   2.853  0.00433 **
## nodefactor.Prison.1     0.21142    0.09384      0   2.253  0.02425 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##      Null Deviance: 1984  on 1431  degrees of freedom
##  Residual Deviance: 1454  on 1425  degrees of freedom
##
## AIC: 1466    BIC: 1497    (Smaller is better.)
```

The model above uses matches on each value for Birthplace. The results show that the homophily effect is seen at all birthplaces except the United Kingdom.

```
##                                     edges triangle
## London Gangs                          315      860
## NULL Model                            292      195
## Birthplace and Prison                 297      260
```

```
## Homophily Birthplace Match          312      266
## Homophily Birthplace Match w/Diff   271      178
```

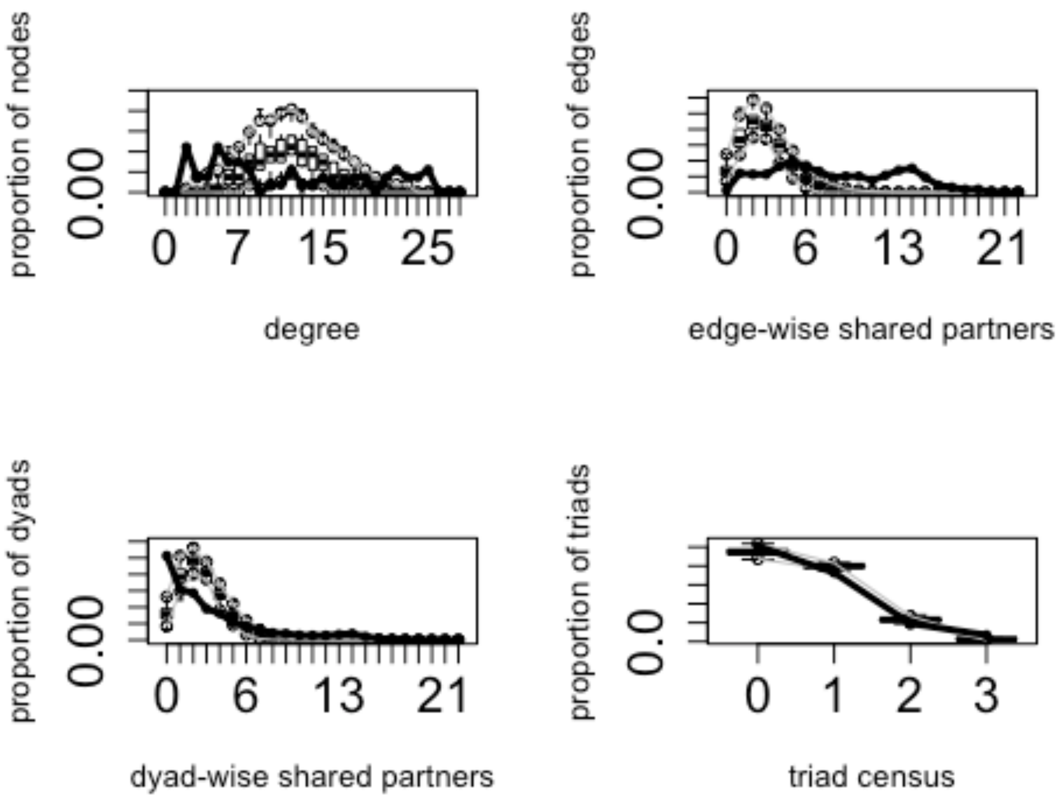
Simulation statistics shows that our homophily models produced mixed results. Homophily with Birthplace match was the closest model on edges and triangles. Again the number of triangles is still significantly lower than the observed network.

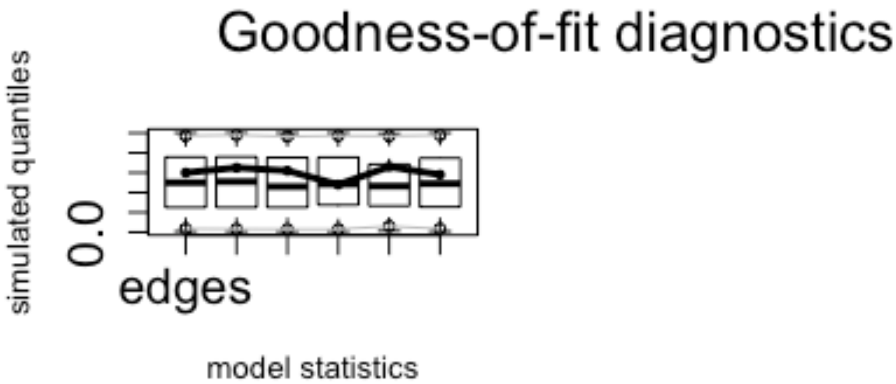
```
##                                         AIC
## NULL Model                         1510.466
## Birthplace and Prison              1472.498
## Homophily Birthplace Match         1502.110
## Homophily Birthplace Match w/Diff 1465.626
```

Analysis of the AIC values show that the Homophily Birthplace Match with Diff model has the best performance (lowest value) out of all the models.

## Goodness of Fit

Using the model with the overall lowest AIC value, Homophily Birthplace Match with Diff, we will compare selected network properties of the simulated networks to those same network characteristics of the observed London Street Gang network. Specifically, we will examine the geodesic distances, the distribution of edgewise shared partners, the degree distribution, and the frequency of different patterns of triangles.


proportion of nodes
0.00
0
7
15
25
degree
proportion of edges
0.00
0
6
13
21
edge-wise shared partners
proportion of dyads
0.00
0
6
13
21
dyad-wise shared partners
proportion of triads
0.0
0
1
2
3
triad census




Examination of the diagnostic plots shows that, aside from the triad census and node attribute Prison = 1, the model struggles to fit the observed network.

## Bibliography

Grund, T. and Densley, J. (2015) Ethnic Homophily and Triad Closure: Mapping Internal Gang Structure Using Exponential Random Graph Models. Journal of Contemporary Criminal Justice, Vol. 31, Issue 3, pp. 354-370

Grund, T. and Densley, J. (2012) Ethnic Heterogeneity in the Activity and Structure of a Black Street Gang. European Journal of Criminology, Vol. 9, Issue 3, pp. 388-406. SOURCE: Available from Manchester.

https://sites.google.com/site/ucinetsoftware/datasets/covert-networks/londongang