

DS6501 Assignment 2

CHUNGUANG LIU 2230873

Executive Summary

This report presents a comprehensive network analysis using two files from “StreetGangData” dataset. The objective of this report is to understand the dynamic interaction between gang members by analyzing the connections, centrality measures, and community structures. To identify key members in the network, degree, betweenness, and closeness centrality measures were calculated based on the gang member igraph object. In the gang member network, nodes represent the gang members and links represent the connection between members/nodes. A simplified network was generated to view the crucial groupings and the influential members. By analyzing the ranking and the birthplace of members who have served in prison, the insights into the interaction patterns are provided, especially among the imprisoned members. The nodes with higher hub score are key connectors. Targeting these members can help the police to prevent the gang activities. Networks between different ethnicities created mainly support the hypothesis that gang members from same ethnicity tend to act together, especially with a detailed analysis of the UK-Caribbean network.

Visual Results & Analysis

Task 1:

After importing the two data file, these two files were checked by using head() method to make sure they are loaded successfully. An igraph object was created to show the network of the London street gang by setting the links as undirected. By using the E(), V(), edge_attr(), and vertex_attr() methods, we can check the nodes, links, and attributes. Figure 1.1 to Figure 1.3 show the details of edges and vertices. There are 54 nodes (gang members) and 315 edges (connections between members). Nodes in this network represent the members of the London street gang. Each vertex or node has several attributes which are detailed information of the gang members, such as name, age, birthplace, residence, arrests, convictions, prison, ranking, and size. Links in the network represent the relationship between the members. Each link or edge has attributes like weight and width.

Figure 1.4 shows the plot with default settings. In terms of readability, there are some problems. First, some of the nodes are overlapping and all nodes are in the same size, which makes it hard to view individual nodes. Second, all edges have the same width, no matter of their weight, which makes it difficult to view stronger relationships between nodes.

The improved plot (see Figure 1.5) set the width of links to the value of their weight. This adjusted the link width making it easier to identify more important links in the network. Also, the size of each node is proportional to its age attribute. Larger nodes represent older members, smaller means younger. This makes it easier to distinguish the nodes based on their age.

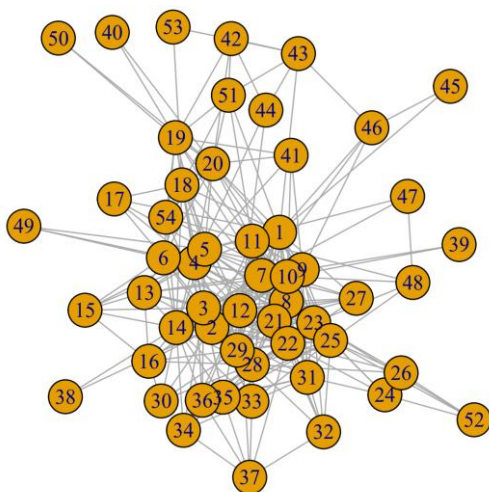


Figure 1.4 Plot with Default Settings

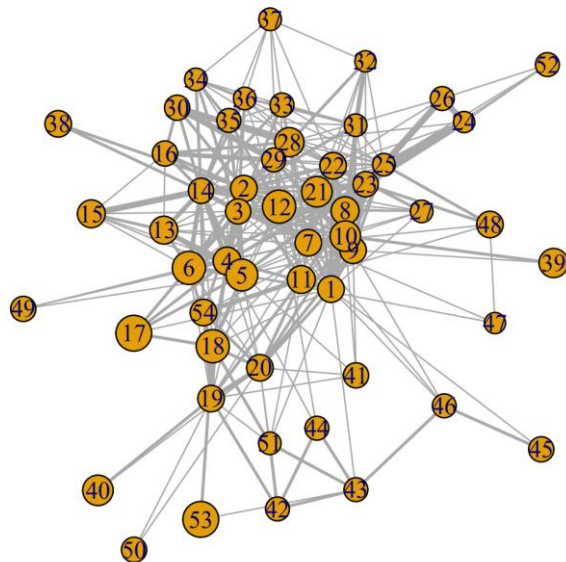


Figure 1.5 Improved Plot with Adjusted Attributes

Task 2:

In this task, degree, betweenness, and closeness were checked. The top 3 nodes with the highest values for each measure have been identified, as shown in Figure 2.1. The main reason to check these centrality measures is to identify who are the most influential gang members. Focusing on these significant members, police can reduce the gang's influence.

Degree is the score of how many of links a node has. Nodes 1, 7, and 12 have the highest degree

centrality (25), which means they have a lot of directed connections with other gang members. They are the most connected and “popular” ones in the network. Targeting these members can let the police disrupt the communication within the network.

Betweenness is calculated based on the number of times a vertex lies on the shortest paths between other nodes. Nodes 1, 7, and 4 have the highest betweenness centrality. They are the crucial bridges in the gang. Targeting these members can influence the flow of the information.

Closeness measures how quickly a node can reach other nodes. Nodes 1, 7, and 12 have the highest closeness centrality, which means they are in the most influential positions. Removing these members can influence the whole gang network.

```
> #(b) descending order of each centrality measure
> #degree
> sorted_degree <- sort(centrality_degree, decreasing = TRUE)
> top3_degree <- head(sorted_degree, 3)
> top3_degree
  1  7 12
25 25 25
> #betweenness
> sorted_betweenness <- sort(centrality_betweenness, decreasing = TRUE)
> top3_betweenness <- head(sorted_betweenness, 3)
> top3_betweenness
      1      7      4
149.74464 104.03338 99.92032
> #closeness
> sorted_closeness <- sort(centrality_closeness, decreasing = TRUE)
> top3_closeness <- head(sorted_closeness, 3)
> top3_closeness
      1      7      12
0.01234568 0.01234568 0.01204819
```

Figure 2.1 Top 3 Nodes with the Highest Values for Degree, Betweenness, and closeness

Task 3:

In this part, a simplified network was created by removing all the nodes whose degree is less than 15 and removing all the edges whose weight is less than 3 (as shown in Figure 3.1). There are three crucial groups and one isolate node in the figure. The first group consists of only two nodes. The second group is moderate in size. The largest group consists of 8 nodes. According to the centrality measures in task 2, the bridge nodes are node 1, 7 and 4. Although node 12 is isolated in this simplified network, its high centrality measures still show it is a significant

connector. Node 1 is likely to be the most influential street gang member, because it has high degree, betweenness, and closeness centrality and it is also a bridge node in its grouping.

simplified network by removing some nodes and edges

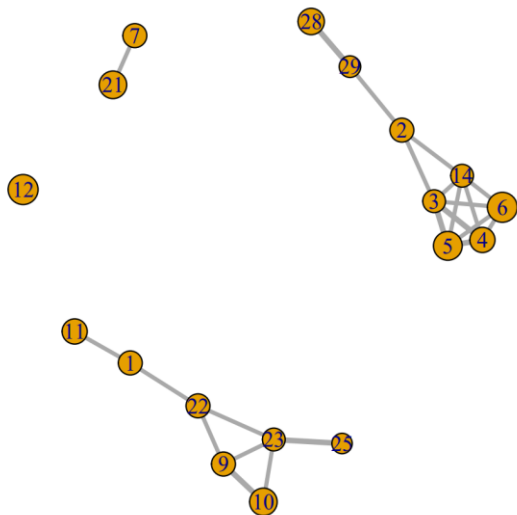


Figure 3.1 Simplified Network by Removing Some Nodes and Edges

Task 4:

Figure 4.1 below shows the network based on node's ranking assigned by police. The value of ranking ranges from 1 to 5, where 1 is the highest displayed in red color, 5 is the lowest ranking shown in light blue. From Figure 4.1, we can see that nodes with same color (same ranking) tend to cluster together. Red, orange, and yellow nodes have higher assigned ranking and are positioned in the center of the network. These higher-ranking members (Ranking 1, 2, and 3) have more connections, meaning they have greater influence and more serious co-offending in gang activities.

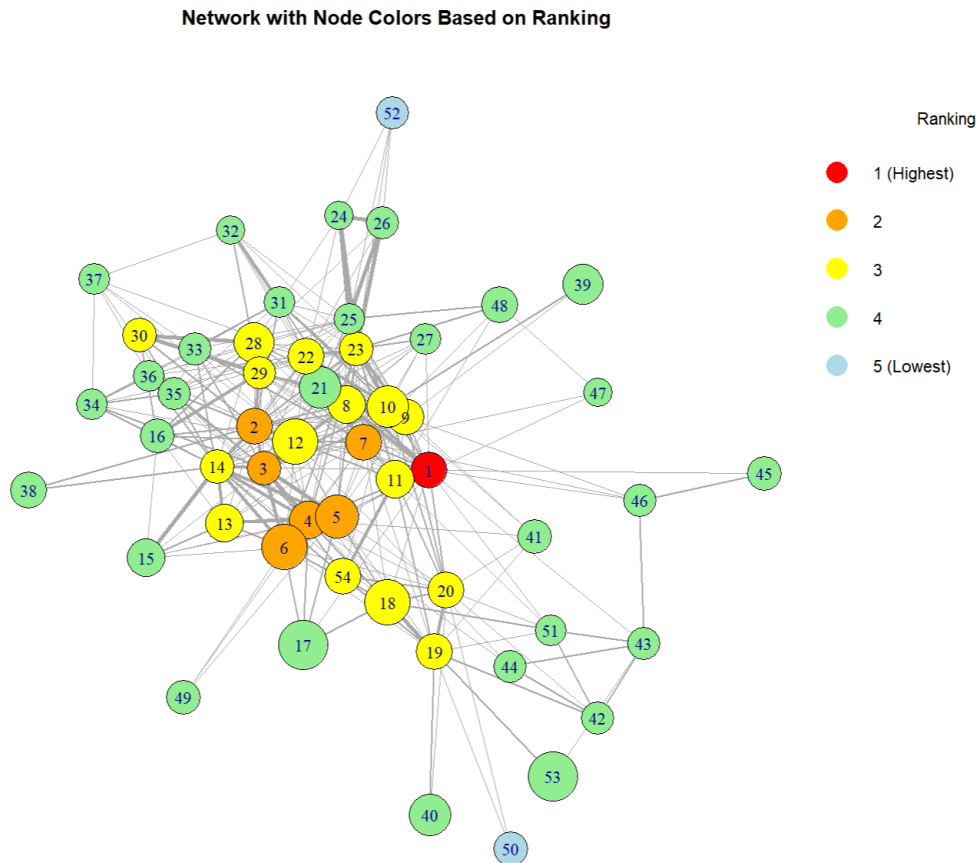


Figure 4.1 Network with Node Colors Based on Ranking

Figure 4.2 was created based on the birthplaces of the gang members who have served time in prison. From this network, there is a cluster of members from West Africa (colored purple), Caribbean (colored cyan), and East African (colored brown). They have more dense connections. In general, members from same ethnicity tends to group together, like the purple West African nodes cluster together. Several Caribbean members act as bridges connecting UK and West/East African groups.

Members Who Have Served Time in Prison in Different Birthplaces

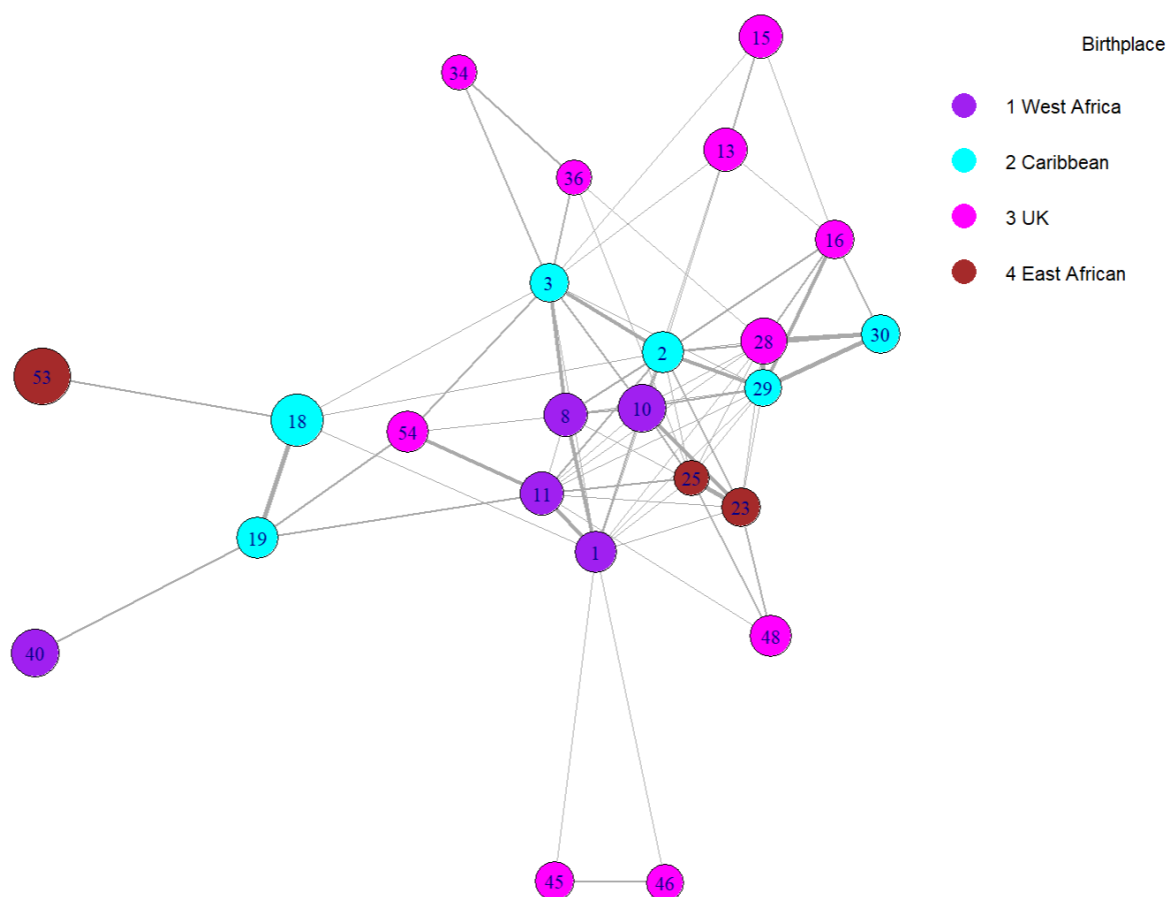


Figure 4.2 Gang Members Who Have Served in Prison in Different Birthplaces

Figure 4.3 is a two-panel plot created basing on Figure 4.2 by removing nodes whose ranking is less than 3. The plot on the left shows the network with nodes size proportional to hub score. The bigger the node the higher the hub score. Nodes with high hub score point to more other important nodes.

There are two communities in the right plot display higher hub score. The first one consists of nodes 16 and 28 from UK (colored magenta) and nodes 29 and 30 from Caribbean (colored cyan). The links in this community is thicker, indicating the interaction among gang members is co-offending and serious. The second community consists of nodes 8 and 10 from West Africa (colored purple) and nodes 23 and 25 from East Africa (colored brown). Node 48 is from UK. Interactions between nodes 10 and 23, and nodes 23 and 25 are more serious. Other

interactions within the community are relatively moderate serious. Node 11 from West Africa plays a crucial role in the second community and another community.

The links between these two communities are dense but not thick, indicating there are communication between them but connections may be not particularly strong.

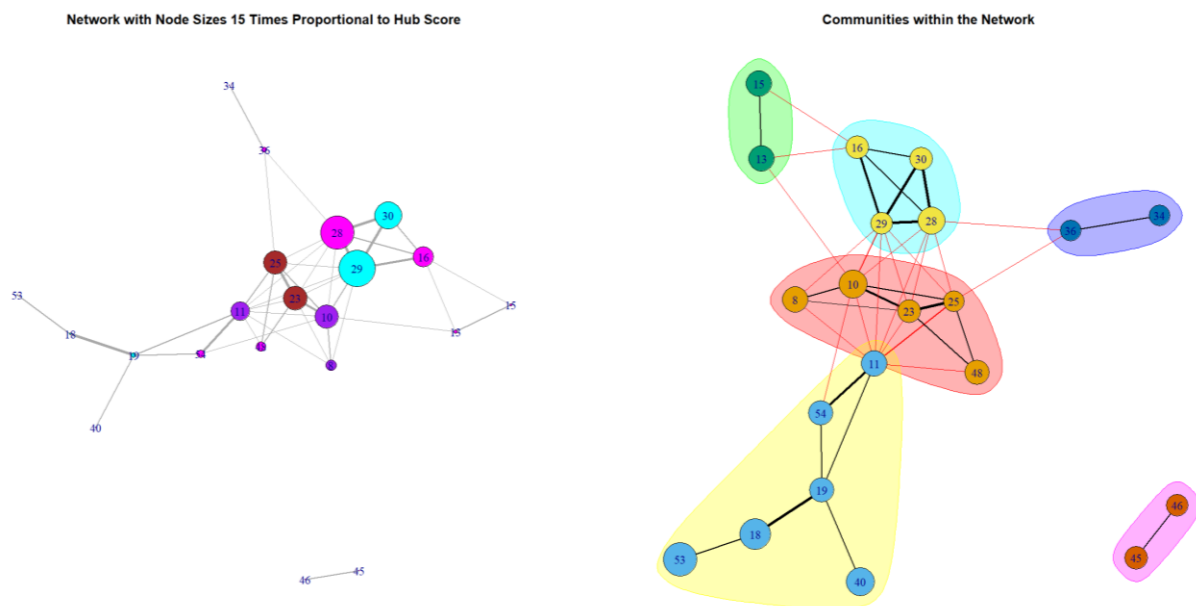


Figure 4.3 A Two-panel Plot Showing Network Visualization with Hub Scores and Community Structures

Task 5:

Figure 5.1 to Figure 5.5 are five networks plotted based on the 'Ranking' attribute. From task 2, we already know that node 1 has the highest degree, betweenness, and closeness centrality. So, even Figure 5.1 does not show dense connection, it is still reasonable to assign node 1 the highest ranking. In Figure 5.2, the node connections are still not dense, but nodes 4 and 7 show high centrality measures in task 2, so it is reasonable to assign ranking 2 to second highest serious. Figure 5.3 shows high node density and high edge connections, ranking 3 is acceptable. Figure 5.4 has the highest node density, but edges connections are relatively low, ranking 4 can show a low ranking in the seriousness of the co-offending. Figure 5.5 only has two nodes and no connections indicating the lowest ranking. In general, the ranking values assigned by police

can reflect the seriousness of the co-offending gang activities among gang members.

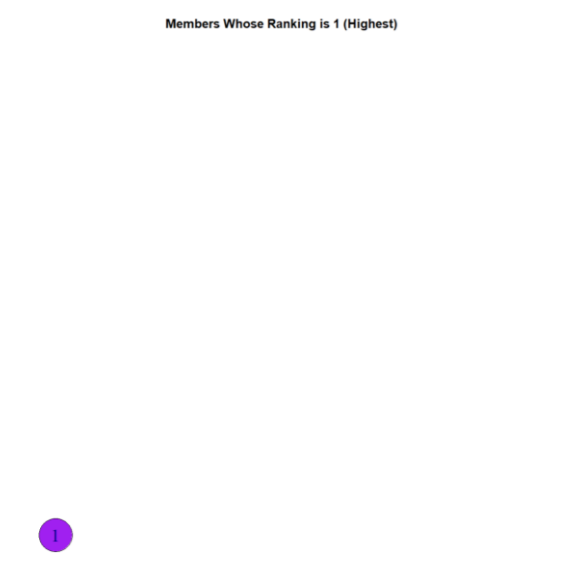


Figure 5.1 Gang Members Ranking 1 (Highest)

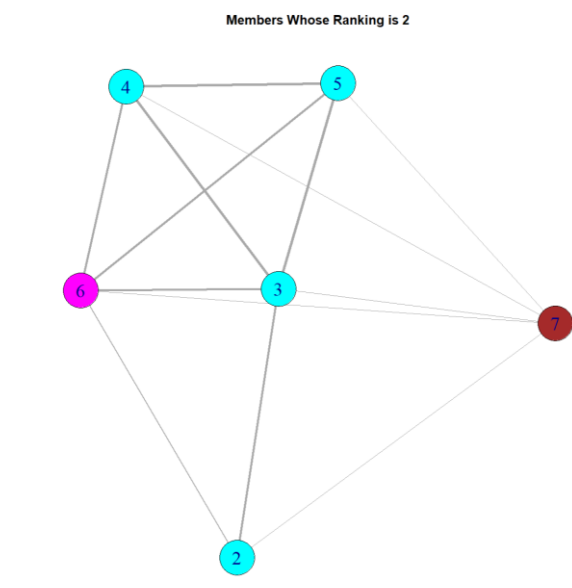


Figure 5.2 Gang Members Ranking 2

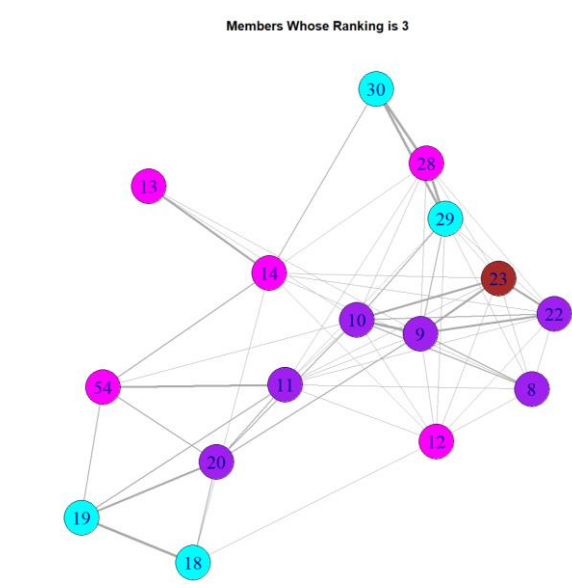


Figure 5.3 Gang Members Ranking 3

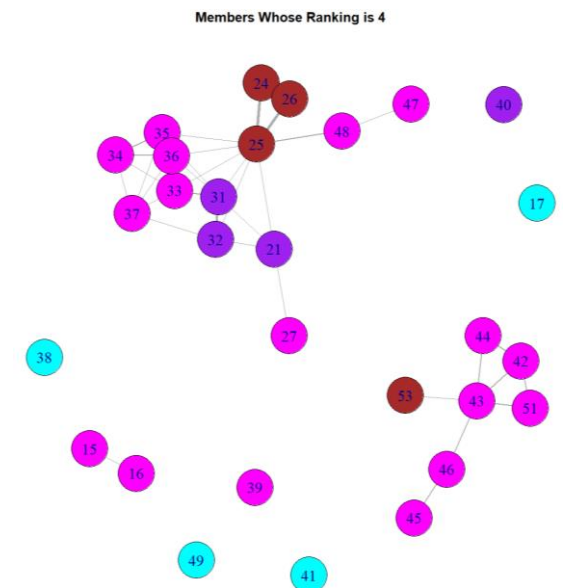


Figure 5.4 Gang Members Ranking 4

52

50

Figure 5.5 Gang Members Ranking 5 (Lowest)

Task 6:

Figures 6.1, 6.2, and 6.3 are networks showing the criminal interactions between UK gang members and members of other ethnic groups (West Africa, Caribbean, and East Africa). Figure 6.4 is a two-panel plots based on the UK-Caribbean criminal interaction network (Figure 6.2), where the node sizes are 10 times the value of authority score in the first panel, and the second panel is the network of community structures. The weight of all the links in these figures are larger than 1, which means that the relationship between nodes are not friendships.

From Figures 6.1, 6.2, and 6.3, we can see that there are still most strong interactions left between UK and other ethnicities, after removing the links whose weight value equals to 1. And the nodes from the same ethnicity tend to cluster together. This can be the evidence to support the hypothesis that gang members mostly co-offend with other members of the same ethnicity. The first panel in Figure 6.4 shows that both two areas have nodes with high authority scores, which can be considered as key figures in the gang. For example, nodes 13, 6, and 14 come from the UK, and nodes 2, 3, 4, and 5 come from Caribbean. These nodes have strong connections, which can support the hypothesis. While in the second panel, it shows that the communities of these high-authority-score nodes are ethnically diverse. For example, nodes 4,

5, 6, and 13 position in the overlap zone of two communities. This can be the evidence to challenge the hypothesis.

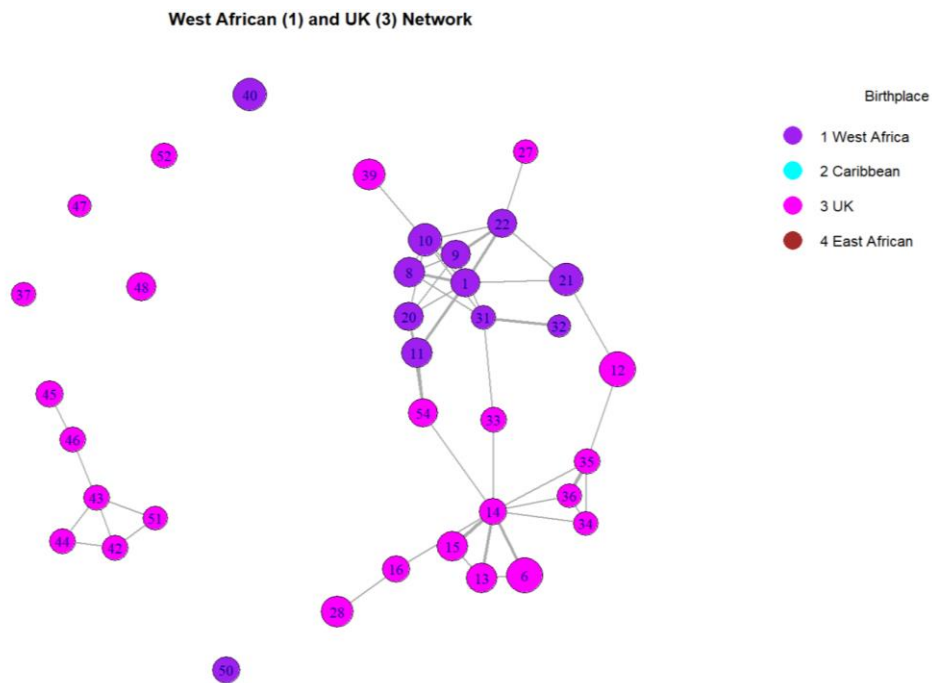


Figure 6.1 Network Showing the Connections between Gang Members from UK and West Africa

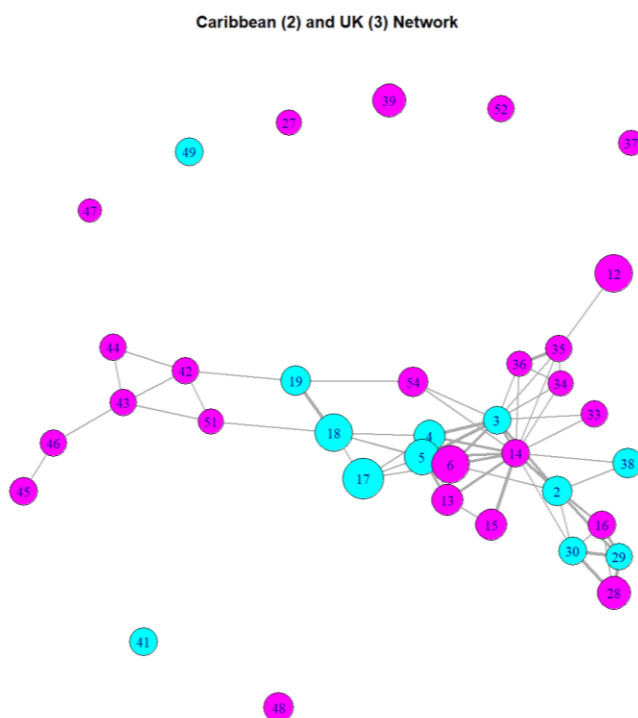


Figure 6.2 Network Showing the Connections between Gang Members from UK and Caribbean

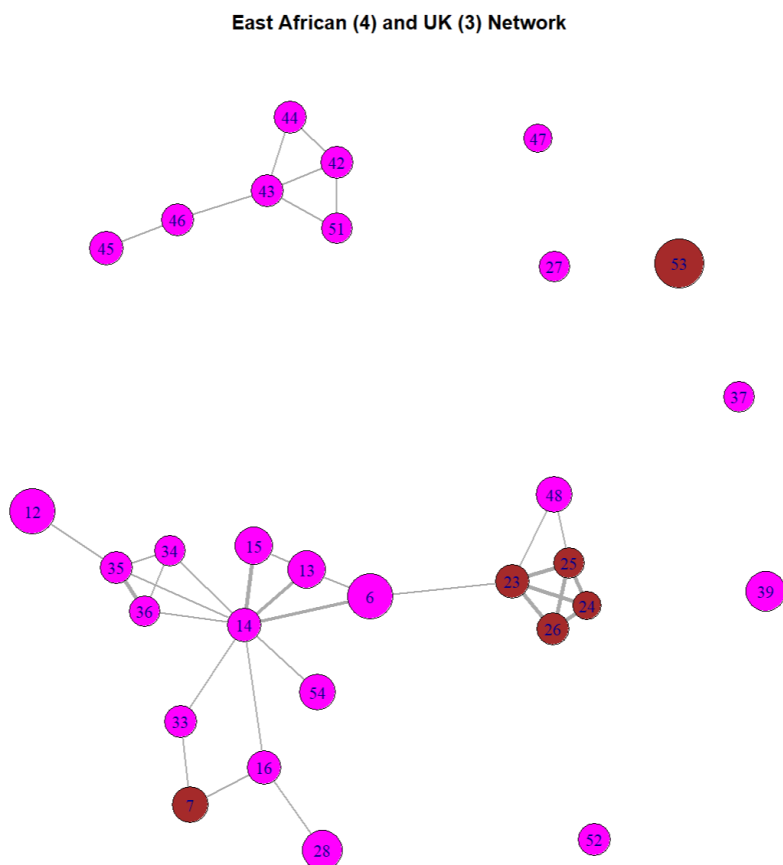


Figure 6.3 Network Showing the Connections between Gang Members from UK and East Africa

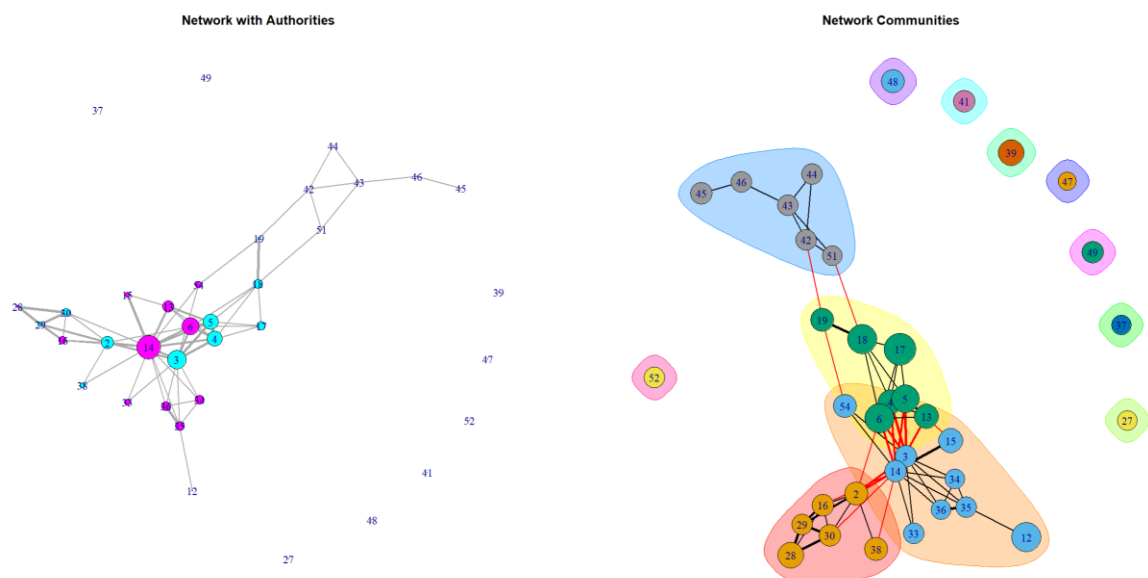


Figure 6.4 A Two-panel Plot Showing Network Visualization with Authority Scores and Community Structures Based on UK-Caribbean Network

Conclusion

The analysis of the London street gang network offers several findings. Nodes with high degrees, betweenness, and closeness centralities, like nodes 1, 7, and 12, are important roles in the gang. Nodes with higher hub scores have higher connections towards other members, like nodes 16, 18, 29, and 30. Nodes with higher authority scores are key figures in the gang, like nodes 13, 6, 14, 3, 4, and 5. Birthplace/ethnicity analysis supports the hypothesis that same ethnic members mainly co-offend. Community detection also support this hypothesis, even though it also shows that high-authority-score nodes are ethnically diverse.

One future research direction can be developing intervention strategies towards the targeted crucial members we analyzed in this report. This can help the police to reduce the gang activities. Another direction can be to analyze and understand how interactions flow between key nodes.