

Project Title

Joana Pimenta, MSc in Artificial Intelligence, 0001168774
Name2 Surname2, Degree Programme, ID number (matricola)

1 Introduction

The present study is situated within the broader field of social sciences, with a particular focus on the application of Social Network Analysis (SNA) to the investigation of covert criminal organizations. Social Network Analysis offers a methodological framework to explore how individuals within a group are connected, how influence and control are distributed, and how structural properties affect organizational resilience and vulnerability. Our specific application concerns the comparative analysis of two covert criminal networks — one operating in Italy and the other in London — in order to explore both structural and sociocultural differences between them. The datasets used for this study, *Italian Gangs Network* (67 nodes, 114 edges) (available online at [1]) and *London Gang Network* (54 nodes, 315 edges) (available online at [2]) describe internal relationships within each gang and share a common metadata attribute: the nationality of each individual.

2 Problem and Motivation

The main objective of this study is to analyze and compare the structural and social dynamics of two covert criminal networks, one Italian and one London-based, in order to understand how relational structure, along with ethnic and national composition, influences internal organization and group resilience. Historically, researchers often overlooked gangs as collective entities, focusing instead on the traits of individual members [3]. As a result, the connection between gang activity and gang structure has remained largely unexplored, described as a “black box” [4]. Nevertheless, a small but expanding literature on gangs and social networks has emerged, based on the idea that “human relationships form the least common denominator for organized crime” [5].

Klein and Maxson [6] highlight that “ethnicity is one of the most widely discussed, and little studied, aspects of gangs.” Ethnic groups—defined by shared heritage, culture, language, religion, or country of birth [7]—play a central role in gang dynamics. Research has consistently shown that ethnicity is a key factor in both gang formation and membership [8] [9] [10]. While ethnic diversity has been recognized as influential in broader community contexts [11], its specific relationship to gang activity and internal organization remains largely unexplored [12].

The project therefore addresses two key issues: first, we seek to understand the structural organization and internal dynamics of covert criminal networks, focusing on how individual positions and relational patterns contribute to the stability and hierarchy of such groups; second, we aim to investigate the sociocultural dimension of these networks, exploring how national composition (i.e., the nationality of members) influences patterns of connection, leadership, and cooperation within and across criminal organizations.

To achieve these aims, the project adopts a Social Network Analysis (SNA) approach, which has a long-standing tradition in gang and organized crime research [13] [14]. Technological advancements have significantly expanded its potential [15] [16]. Reconstructing a gang as a social network involves linking each unit—whether a group or an individual—according to the type of relationship under investigation [17]. This dual approach is essential not only to advance the theoretical application of Social Network Analysis within criminology, but also to generate actionable insights for security policies and investigative strategies by identifying key nodes and structural vulnerabilities in criminal systems. Although prior research has examined individual network structures, comparative studies across distinct sociocultural contexts remain rare. From a practical perspective, understanding how relational configurations and cultural factors interact to influence organizational resilience can inform policymaking, intelligence analysis, and law enforcement interventions, facilitating the detection of central actors, weak points, and cohesive subgroups that sustain covert operations.

The main contributions of the project include:

1. Comparative Perspective - We will conduct a systematic comparison between two real-world covert criminal networks (Italian and London-based), highlighting structural and social differences.
2. Integration of Structural and Sociocultural Analysis - By combining network metrics with the nationality attribute, we aim to understand how social identity and homophily affect

organizational patterns and leadership roles.

3. 3. Insights into Network Robustness - Through the evaluation of cohesion and vulnerability to the removal of central nodes, the project will provide a deeper understanding of how criminal organizations maintain resilience despite potential disruptions.

This project aims to contribute to the understanding of relational and cultural dynamics that shape criminal networks, providing both a theoretical foundation and methodological tools for future research and practical applications in the field of security and crime prevention.

3 Datasets

3.1 Gang Network Datasets and Processing

Our analysis uses two publicly available datasets from the UCINET Software project’s repository of social networks: the **London Gang Network** (54 nodes) and the **Italian Gangs Network** (67 nodes). The London dataset was accessed from the UCINET “Covert Networks” collection [2]; the Italian dataset was obtained from the same repository [1]. Both include person–person adjacency matrices and accompanying node-attribute tables, and are fully digitised, requiring no manual transcription.

3.1.1 Digitisation and Data Handling

For this study, each adjacency matrix was stored in CSV format (`LONDON_GANG.csv` with a 54×54 matrix; `ITALIAN_GANGS.csv` with a 67×67 matrix). The London attributes file (`LONDON_GANG_ATTR.csv`) lists Age, Birthplace, Residence, Arrests, Convictions, Prison, Music, and Ranking; the Italian attributes file (`ITALIAN_GANGS_ATTR.csv`) includes Nationality/Country of origin. To enable cross-network comparisons, we used only a harmonised birthplace variable (London Birthplace; Italian Nationality/Country of origin). Data handling was performed in **Python** with `pandas`, using `pandas.read_csv` to load matrices and attributes into `DataFrames`.

3.1.2 Computing Measures

Network measures were computed with `NetworkX`. The adjacency `DataFrames` were converted into graph objects via `nx.from_pandas_adjacency`, which then served as the basis for all subsequent analyses and visualisations.

4 Validity and Reliability

How closely does the model of your dataset represent reality (validity)? How does the way you treat the data affect the reproducibility of the study (reliability)?

4.1 London Gang Network Dataset

4.1.1 Validity (Representation of Reality)

The model of the dataset—a graph of 54 nodes and 315 edges—is a structural abstraction of a complex, real-world social system. The validity, or how closely this model represents reality, is subject to several key considerations:

- **Incompleteness of Covert Data:** The dataset maps a "covert network." By definition, such networks are hidden. The data (likely sourced from surveillance or police records) is almost certainly an incomplete snapshot. We must assume that some real-world relationships were unobserved and are missing from the model.
- **Static vs. Dynamic Reality:** The dataset represents the network at a single point in time. Real-world social structures are dynamic, with ties forming, dissolving, and changing in strength. Our model does not capture this temporal evolution.
- **Unweighted Analysis of Weighted Data:** The source data is weighted (with values such as 1, 2, and 3), likely representing the frequency or strength of the relationship. In our analysis, we employed standard, unweighted measures (e.g., `nx.density`, `nx.diameter`, `nx.degree_centrality`). This was an intentional choice to focus purely on the **topological structure**, but it is a significant simplification. The model treats a strong, frequent bond as equivalent to a weak, infrequent one, which impacts the real-world interpretation of influence and cohesion.

In summary, the model is a valid (as it is academically vetted) but simplified, static, and unweighted representation of the network's topology, not a complete or dynamic reflection of its real-world social complexity.

4.1.2 Reliability (Reproducibility)

The reliability of this study—the ability for another researcher to reproduce the exact same results—is **high**. This is ensured by the methodology used to treat the data:

- **Public Data:** The dataset was sourced from a stable, public, and citable URL. Any researcher can access the exact same source file (`LONDON_GANG.csv`), eliminating data collection as a variable.
- **Open-Source, Deterministic Tools:** The entire analysis was conducted using open-source Python libraries (`pandas` and `NetworkX`). The functions used for calculating measures (`nx.density`, `nx.betweenness_centrality`, etc.) are deterministic. Given the same input graph, they will produce the identical output every time.
- **Transparent Workflow:** The data treatment was minimal and explicit: loading the CSV via `pandas`, handling indices, converting it to a `NetworkX` graph, and applying specific functions. This step-by-step process can be scripted and shared, ensuring perfect reproducibility.

5 Measures and Results

MAYBE WE SHOULD INCLUDE THE MATHEMATICAL FORMULAS OF THE METRICS

For both the Italian and London gang, we study the same metrics and compare them. In fact, we start by studying **general structural metrics**. These include:

- density which is the ratio between existing ties and all possible ties and measures cohesion (high density means that communication is easier and that there is lower vulnerability to central node removal);
- average degree which is defined as the average number of connections per node and indicates member activity and level of engagement;
- network diameter and average path length which are respectively the maximum and average length of paths between nodes and are able to measure the network's efficiency in transmitting information or orders;
- clustering coefficient which corresponds to the probability that a node's neighbors are connected to each other and highlights closed subgroups or internal "cells", it is useful for understanding resilience and community;
- modularity which measures the presence of well-defined internal communities and can reveal internal divisions and possible subgroups or cliques.

Besides, we study **centrality metrics**. In particular, we focus on:

- degree centrality which is based on the number of direct connections a node has and is useful to identify the most active or influential members of the network;
- betweenness centrality which is the number of times a node lies on the shortest paths between other nodes and highlights brokers or gatekeepers (nodes critical for the flow of information);
- closeness centrality which is defined as the reciprocal of the sum of a node's distances to all other nodes. A node close to all others can quickly spread information or orders;
- eigenvector centrality. This metric defines importance by being connected to other important nodes and highlights leaders recognized by the most influential members.

Finally, we study the roles and vulnerability based on the metrics calculated previously we aim to identify key roles like leader, broker and peripheral members. We use a combination of centrality, degree, betweenness and clustering to identify who in the network leads, mediates between subgroups and remains peripheral.

We define leaders as the nodes that have both degree and eigenvector centrality values among the top 5% of all nodes: leaders are both broadly connected (high degree) and well positioned near other influential nodes (high eigenvector). Brokers are those that fall within the top 5% for betweenness centrality, meaning they often act as bridges between different parts of the network. Peripherals are nodes in the bottom 5% for degree centrality, indicating that they have few connections and limited influence within the network.

We also study the k-core and core-periphery structure to inspect implicit hierarchy and concentration of power. A k-core is the part of the network where every node has at least k connections to other nodes within that part. We computed core numbers for all nodes and identified the main core - the highest k that still has nodes. The main core has k equal to the maximum core number in the network, and we report how many nodes it contains (out of the total) and list which nodes belong to it.”

We study the cohesion and network robustness by studying metrics like density, average path length and k-core decomposition in a network where the central nodes have been removed and compare them to the values of the original network to study the impact of removing central nodes.

At last, in order to explore how ethnic background influences the internal structure of both the Italian and London gang networks, we conduct a series of **attribute-based network analyses** focusing on the *Birthplace* variable.

The objective is to determine whether individuals tend to associate mainly with others of the same national origin ethnic homophily or whether the networks exhibit patterns of cross-ethnic integration.

We use the following key-metrics:

- **Assortativity coefficient and mixing matrix:** to quantify the extent of homophily by *Birthplace* and to visualize how frequently connections occur within or between different nationality groups.
- **Centrality measures (degree, betweenness, eigenvector):** to identify whether specific ethnic groups occupy more central or influential structural positions in the network.
- **Community composition and diversity:** communities were detected using a modularity-based algorithm, and their ethnic heterogeneity was assessed through the **Shannon Diversity Index (H)**, which measures the balance and variety of nationalities within each community. To quantify this diversity, the Shannon index was computed as:

$$H = -\sum_i p_i \log(p_i)$$

where p_i represents the proportion of members from group i within a community. Higher H values indicate more ethnically diverse communities.

- **Inter-group connectivity:** calculated as the proportion of edges linking individuals of different *Birthplace* categories, indicating the level of cross-ethnic interaction.
- **Subgraph analysis by Birthplace:** used to evaluate the internal cohesion of each nationality group in terms of network density and clustering coefficient.

It is worth noting that the Italian network is not connected, and therefore the network diameter and average path length are calculated for the largest connected component of the network. This is also true when we remove the central nodes.

5.1 Italian gang

5.1.1 Structural analysis

Using the Library Networkx [18] from Python we studied different aspects of the Italian gang network.

In figure 1 a plot of the network is shown where each node is colored according to its country label.

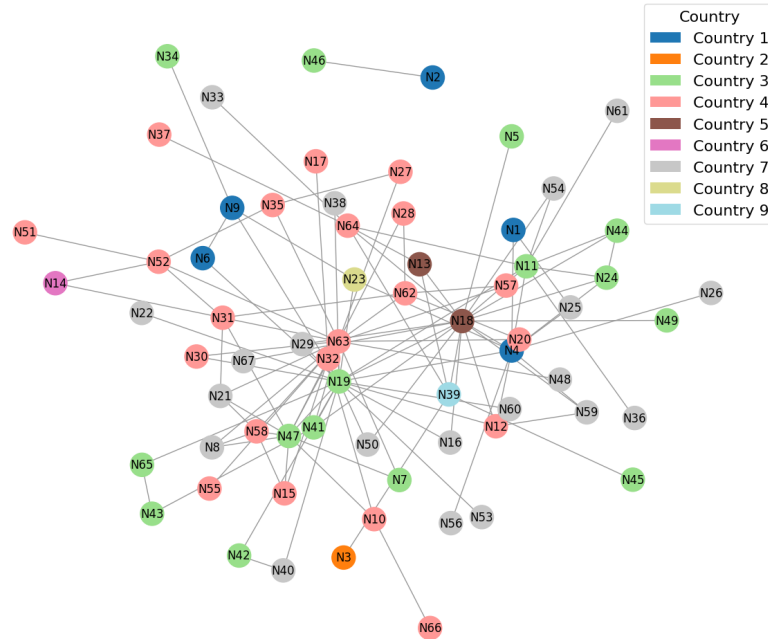


Figure 1: Italian Network graph visualization. Each node is colored according to its country label

5.1.2 Macro-level Cohesion and Structure

| Metric | Result | Interpretation (What it means) |
|------------------------|--------|---|
| Density | 0.0516 | About 5% of all possible connections exist, so the network is fairly sparse. |
| Average Degree | 3.0430 | On average, each member is connected to 3 others. |
| Average Path Length | 3.012 | Any two members can reach each other in about 3 "hops" on average. |
| Diameter | 6 | The maximum separation between any two members is 6 "hops". |
| Avg. Clustering Coeff. | 0.4347 | Members' friends are often friends with each other—around 43% of possible "triangles" are closed—showing moderate local cohesion. |
| Modularity | 0.5561 | The network splits into well-defined communities; there are many more connections within groups than between them. |

5.1.3 Micro-level Centrality and Social Roles

These measures identify the most important nodes, allowing us to define social roles.

Leaders The most influential, connected, and central members.

| Node | Degree | Betweenness | Closeness | Eigenvector |
|------|--------|-------------|-----------|-------------|
| 19 | 0.3182 | 0.5558 | 0.5397 | 0.4394 |
| 63 | 0.2879 | 0.3633 | 0.4597 | 0.3110 |
| 18 | 0.2727 | 0.2881 | 0.4563 | 0.3784 |

Table 1: Leader nodes (Degree ≥ 0.1621 & Eigenvector ≥ 0.2542)

Brokers Members who act as "bridges" connecting different parts of the network.

| Node | Degree | Betweenness | Closeness | Eigenvector |
|------|--------|-------------|-----------|-------------|
| 19 | 0.3182 | 0.5558 | 0.5397 | 0.4394 |
| 63 | 0.2879 | 0.3633 | 0.4597 | 0.3110 |
| 18 | 0.2727 | 0.2881 | 0.4563 | 0.3784 |
| 47 | 0.1515 | 0.1396 | 0.4280 | 0.2177 |

Table 2: Broker nodes (Betweenness ≥ 0.1369)

Peripheral Members Marginal members with few connections, existing on the network's edges.

| Node | Degree | Betweenness | Closeness | Eigenvector |
|------|--------|-------------|-----------|-------------|
| 2 | 0.0152 | 0.0000 | 0.0000 | |
| 3 | 0.0152 | 0.0132 | 0.0000 | |
| 5 | 0.0152 | 0.0556 | 0.0000 | |
| 17 | 0.0152 | 0.0141 | 0.0000 | |
| 22 | 0.0152 | 0.0645 | 0.0000 | |
| 26 | 0.0152 | 0.0384 | 0.0000 | |
| 33 | 0.0152 | 0.0179 | 0.0000 | |
| 34 | 0.0152 | 0.0140 | 0.0000 | |
| 36 | 0.0152 | 0.0349 | 0.0000 | |
| 37 | 0.0152 | 0.0179 | 0.0000 | |
| 38 | 0.0152 | 0.0645 | 0.0000 | |
| 45 | 0.0152 | 0.0132 | 0.0000 | |
| 46 | 0.0152 | 0.0000 | 0.0000 | |
| 48 | 0.0152 | 0.0457 | 0.0000 | |
| 49 | 0.0152 | 0.0556 | 0.0000 | |
| 51 | 0.0152 | 0.0117 | 0.0000 | |
| 53 | 0.0152 | 0.0645 | 0.0000 | |
| 55 | 0.0152 | 0.0457 | 0.0000 | |
| 56 | 0.0152 | 0.0384 | 0.0000 | |
| 60 | 0.0152 | 0.0645 | 0.0000 | |
| 61 | 0.0152 | 0.0349 | 0.0000 | |
| 66 | 0.0152 | 0.0069 | 0.0000 | |
| 67 | 0.0152 | 0.0556 | 0.0000 | |

Table 3: Peripheral nodes (Degree ≤ 0.0152)

5.1.4 Hierarchy and Vulnerability

These measures test the power structure and resilience of the network.

K-Core Decomposition The most densely connected "core" of the network is identified. The analysis reveals a main core with a **k-value of 3**. This core consists of **20 nodes** (out of 63). The nodes in this core are: [4, 8, 11, 12, 13, 15, 18, 19, 21, 24, 31, 32, 39, 41, 44, 47, 58, 59, 63, 64].

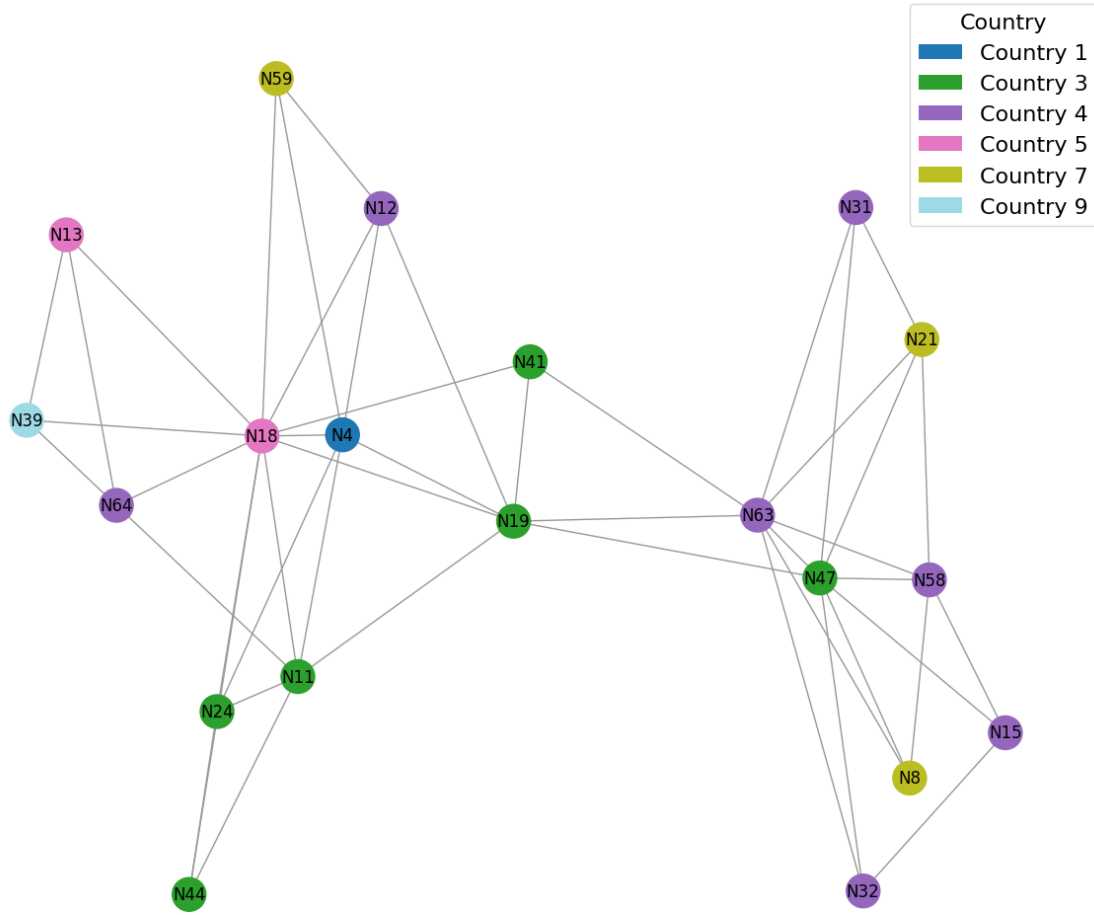


Figure 2: Core of Italian Network graph visualization.

Vulnerability Simulation We simulated the removal of the identified leaders (Nodes 19, 63, 18), and recalculated cohesion metrics. This measure connects roles (leaders) to cohesion (robustness). The network density was reduced to **0.1014**, and the average path length increased to **3.1341** (an **4.05%** increase), indicating marginally longer communication chains and reduced efficiency.

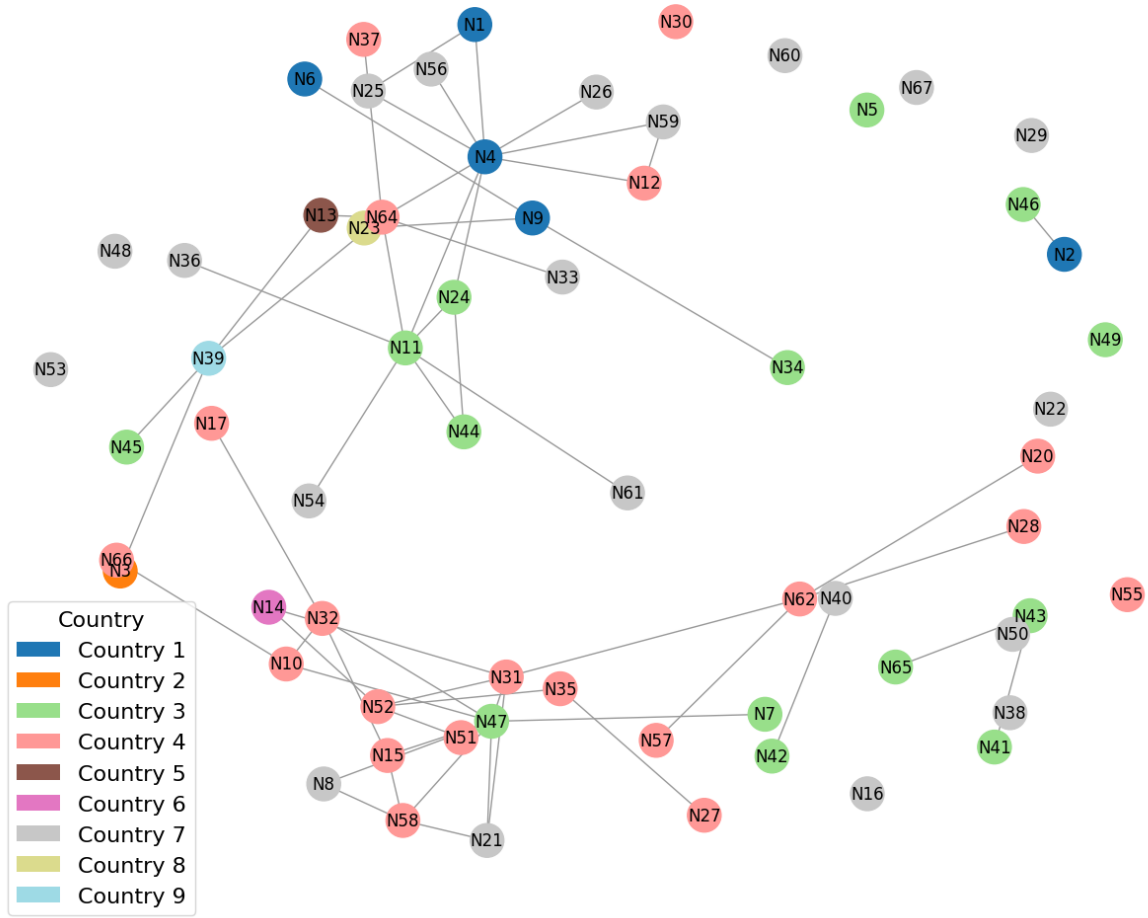


Figure 3: Italian Network sub-graph visualization without leaders.

5.1.5 Synthesis: Connection Between Data, Measures, and Properties

The **Italian gang network** is sparse but moderately cohesive, with clear community divisions and a defined inner core ($k = 3$). Leadership is concentrated in a few nodes (19, 63, 18), supported by brokers who link subgroups. When these leaders are removed, the network becomes **fragmented**, revealing limited redundancy and a **moderate structural vulnerability** despite its local cohesion.

5.1.6 Ethnicity analysis

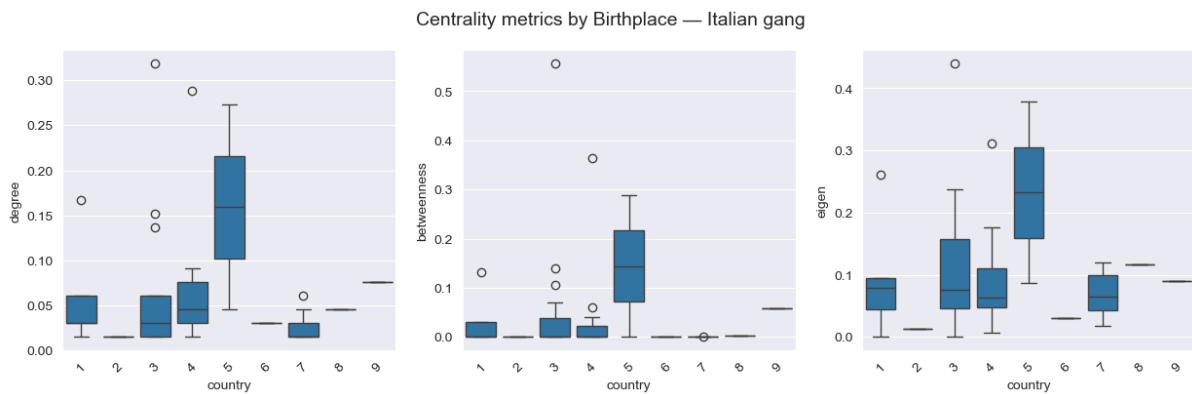
The analysis of the Italian gang network shows a **moderate tendency toward ethnic homophily**, with an assortativity coefficient of 0.150. This value suggests that individuals show a slight preference for connecting with others sharing the same *Birthplace*, though overall the network remains relatively integrated.

The **mixing matrix** confirms this trend: while several diagonal values (e.g., for groups 3, 4, and 5) are higher (indicating intra-group cohesion) many off-diagonal entries are also non-negligible. This demonstrates a considerable number of **cross-national connections**, supporting the idea of partial ethnic mixing.

Table 4: Mean centrality by Birthplace – Italian gang

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 1 | 0.018 | 0.000 | 0.031 | 0.004 | 0.004 | 0.000 | 0.022 | 0.009 | 0.000 |
| 2 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.004 |
| 3 | 0.031 | 0.000 | 0.096 | 0.053 | 0.031 | 0.000 | 0.061 | 0.004 | 0.004 |
| 4 | 0.004 | 0.000 | 0.053 | 0.228 | 0.013 | 0.009 | 0.035 | 0.000 | 0.004 |
| 5 | 0.004 | 0.000 | 0.031 | 0.013 | 0.009 | 0.000 | 0.026 | 0.000 | 0.009 |
| 6 | 0.000 | 0.000 | 0.000 | 0.009 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 7 | 0.022 | 0.000 | 0.061 | 0.035 | 0.026 | 0.000 | 0.000 | 0.000 | 0.000 |
| 8 | 0.009 | 0.000 | 0.004 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 9 | 0.000 | 0.004 | 0.004 | 0.004 | 0.009 | 0.000 | 0.000 | 0.000 | 0.000 |

When analyzing **centrality measures**, group 5 clearly stands out as the most central and structurally influential, with the highest mean degree (0.159), betweenness (0.144), and eigenvector centrality (0.233). Groups 3 and 9 also exhibit moderate centrality levels, suggesting participation in brokerage or connective roles. In contrast, groups such as 2, 6, and 7 show minimal centrality values, occupying peripheral positions within the network. Overall, influence appears somewhat concentrated but not monopolized by a single nationality.



The **community analysis** identified five major communities. Most of these display mixed ethnic compositions, with only one cluster (community 4) being entirely dominated by two groups (1 and 3). The mean Shannon diversity index ($H = 1.174$) indicates high internal heterogeneity, suggesting that communities are composed of members from multiple national origins rather than segregated along ethnic lines.

Furthermore, 64.91% of all connections occur between individuals of different Birthplace categories, a clear indicator of **strong cross-ethnic integration**. This finding supports the interpretation that ethnic background is not a dominant organizing factor in the structure of the Italian gang.

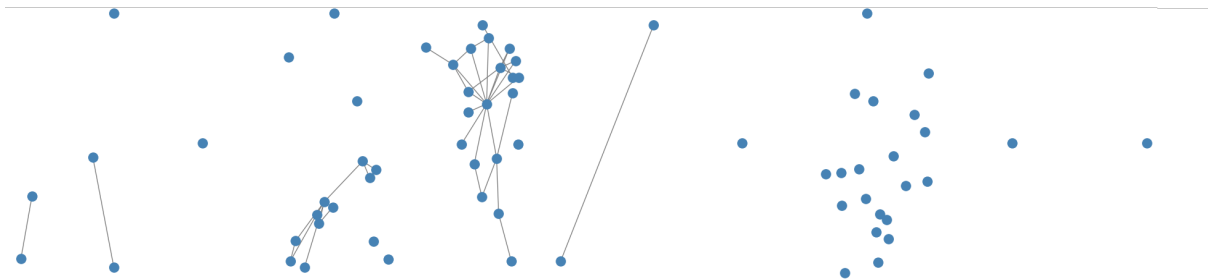
Subgraph analysis by *Birthplace* reveals additional information:

- Group 4 (21 nodes, density = 0.124) forms the largest and most internally cohesive subcommunity.

- Group **5**, although small (2 nodes), is fully interconnected (density = 1.000), representing a tightly bonded dyad.
- Other groups (6, 7, 8, 9) show limited or no internal links, implying **dependence on inter-ethnic ties** for maintaining connectivity.

Table 5: Subgraph-level statistics by Birthplace – Italian gang

| Country | Nodes | Edges | Density | Clustering |
|---------|-------|-------|---------|------------|
| 1 | 5 | 2 | 0.200 | 0.000 |
| 2 | 1 | 0 | 0.000 | 0.000 |
| 3 | 15 | 11 | 0.105 | 0.156 |
| 4 | 21 | 26 | 0.124 | 0.294 |
| 5 | 2 | 1 | 1.000 | 0.000 |
| 6 | 1 | 0 | 0.000 | 0.000 |
| 7 | 20 | 0 | 0.000 | 0.000 |
| 8 | 1 | 0 | 0.000 | 0.000 |
| 9 | 1 | 0 | 0.000 | 0.000 |



In summary, the Italian gang exhibits **moderate homophily but high overall integration**, with collaboration patterns largely transcending national divisions. Ethnicity appears to play a **secondary role** in shaping relational dynamics.

5.2 London gang

This section will report the results of the metrics mentioned in 5 regarding the London gang network 54 nodes, 315 edges.

In figure 4 a plot of the network is shown where each node is colored according to its birthplace label.

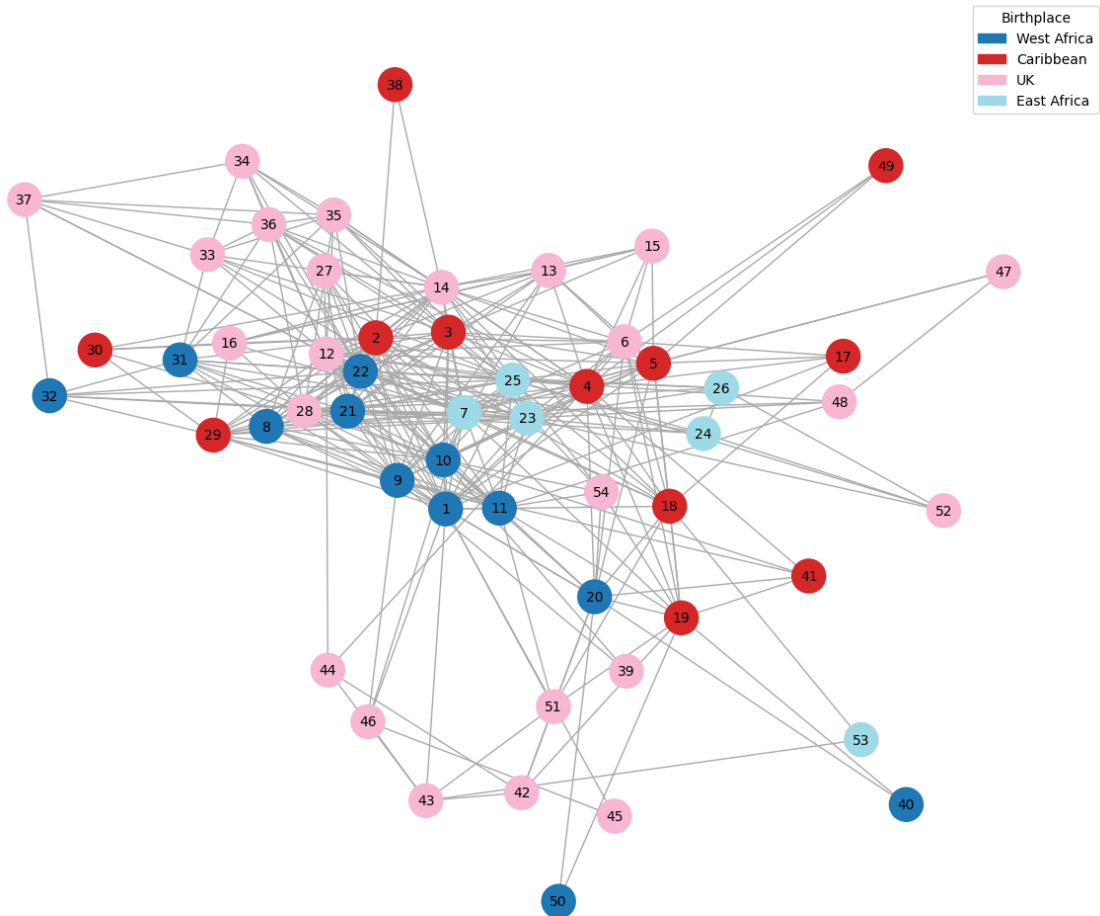


Figure 4: London Network graph visualization. Each node is colored according to its birthplace label

5.2.1 Macro-level Cohesion and Structure

These measures assess the overall "connectedness" and efficiency of the network as a whole.

| Metric | Result | Interpretation (What it means) |
|------------------------|---------------|---|
| Density | 0.2201 | The network is extremely dense and highly interconnected. |
| Average Degree | 11.67 | On average, each member is connected to almost 12 others. |
| Average Path Length | 2.05 | Any two members can reach each other in just 2 "hops" on average. |
| Diameter | 4 | The maximum separation between any two members is 4 "hops". |
| Avg. Clustering Coeff. | 0.6331 | The network is rich in tightly-knit local subgroups (cliques). |
| Modularity | 0.2665 | The network operates as a single, cohesive bloc; it is not fragmented into separate factions. |

5.2.2 Micro-level Centrality and Social Roles

These measures identify the most important nodes, allowing us to define social roles.

Leaders The most influential, connected, and central members.

| Node | Degree | Betweenness | Closeness | Eigenvector |
|-------------|---------------|--------------------|------------------|--------------------|
| 1 | 0.4717 | 0.1087 | 0.6543 | 0.2367 |
| 7 | 0.4717 | 0.0755 | 0.6543 | 0.2433 |
| 12 | 0.4717 | 0.0596 | 0.6386 | 0.2494 |

Table 6: Leader nodes (Degree ≥ 0.4594 & Eigenvector ≥ 0.2357)

Brokers Members who act as "bridges" connecting different parts of the network.

| Node | Degree | Betweenness | Closeness | Eigenvector |
|-------------|---------------|--------------------|------------------|--------------------|
| 1 | 0.4717 | 0.1087 | 0.6543 | 0.2367 |
| 7 | 0.4717 | 0.0755 | 0.6543 | 0.2433 |
| 4 | 0.3962 | 0.0725 | 0.6163 | 0.1747 |

Table 7: Broker nodes (Betweenness ≥ 0.0670)

Peripheral Members Marginal members with few connections, existing on the network's edges.

| Node | Degree | Betweenness | Closeness | Eigenvector |
|-------------|---------------|--------------------|------------------|--------------------|
| 38 | 0.0377 | 0.0000 | 0.4109 | 0.0256 |
| 39 | 0.0377 | 0.0000 | 0.3926 | 0.0258 |
| 40 | 0.0377 | 0.0000 | 0.3557 | 0.0107 |
| 45 | 0.0377 | 0.0000 | 0.4015 | 0.0164 |
| 50 | 0.0377 | 0.0000 | 0.3557 | 0.0107 |
| 53 | 0.0377 | 0.0003 | 0.3681 | 0.0088 |

Table 8: Peripheral nodes (Degree ≤ 0.0377)

5.2.3 Hierarchy and Vulnerability

These measures test the power structure and resilience of the network.

K-Core Decomposition The most densely connected "core" of the network is identified. The analysis reveals a main core with a **k-value of 11**. This core consists of **13 nodes** (out of 54). The nodes in this core are: [1, 2, 7, 8, 9, 10, 11, 12, 21, 22, 23, 25, 29].

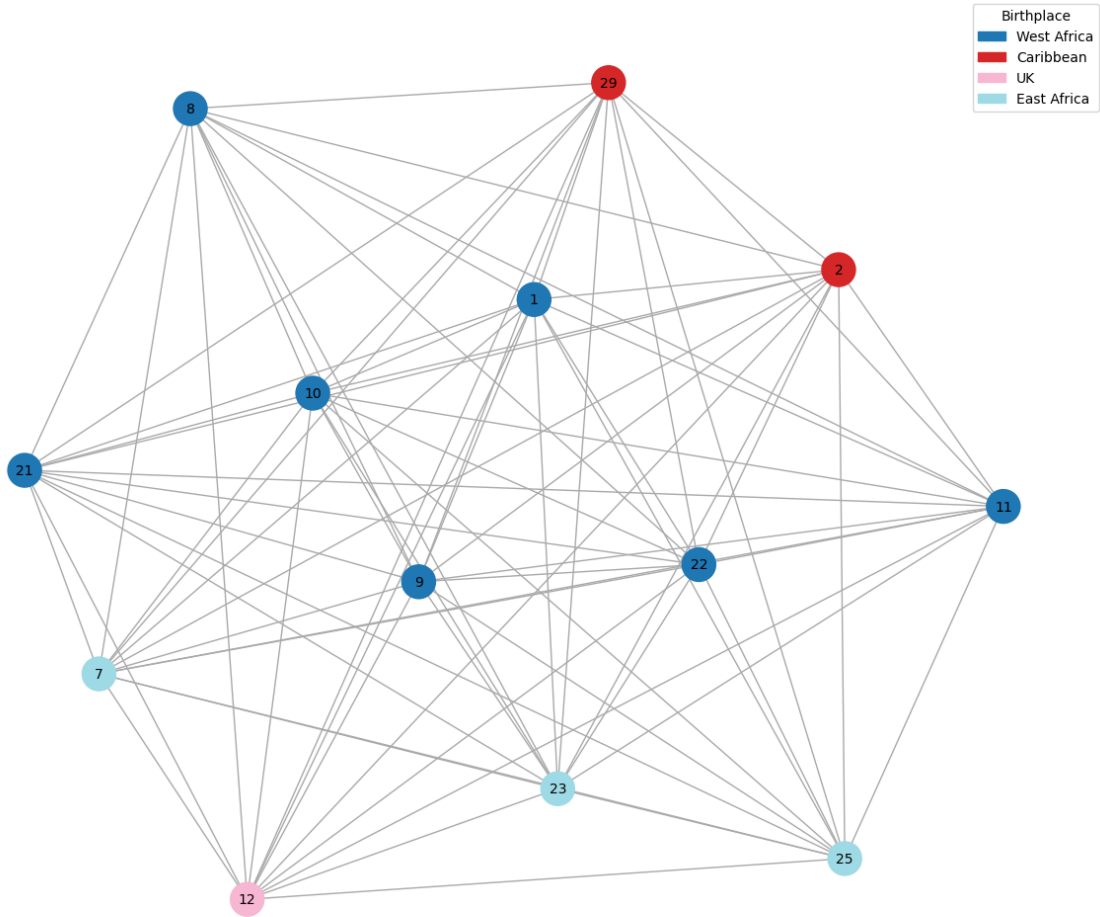


Figure 5: Core of London Network graph visualization.

Vulnerability Simulation We simulated the removal of the identified leaders (Nodes 1, 7, 12), representing a **5.56% reduction in nodes**, and recalculated cohesion metrics. This measure connects roles (leaders) to cohesion (robustness). The results show the network is **extremely robust**: removing the top 3 leaders did not fragment the network (it remained connected). The network density was reduced to **0.1906**, and the average path length increased to **2.2180** (an **8.00%** increase). The high density and clustering create redundancy, making the network resilient to targeted attacks.

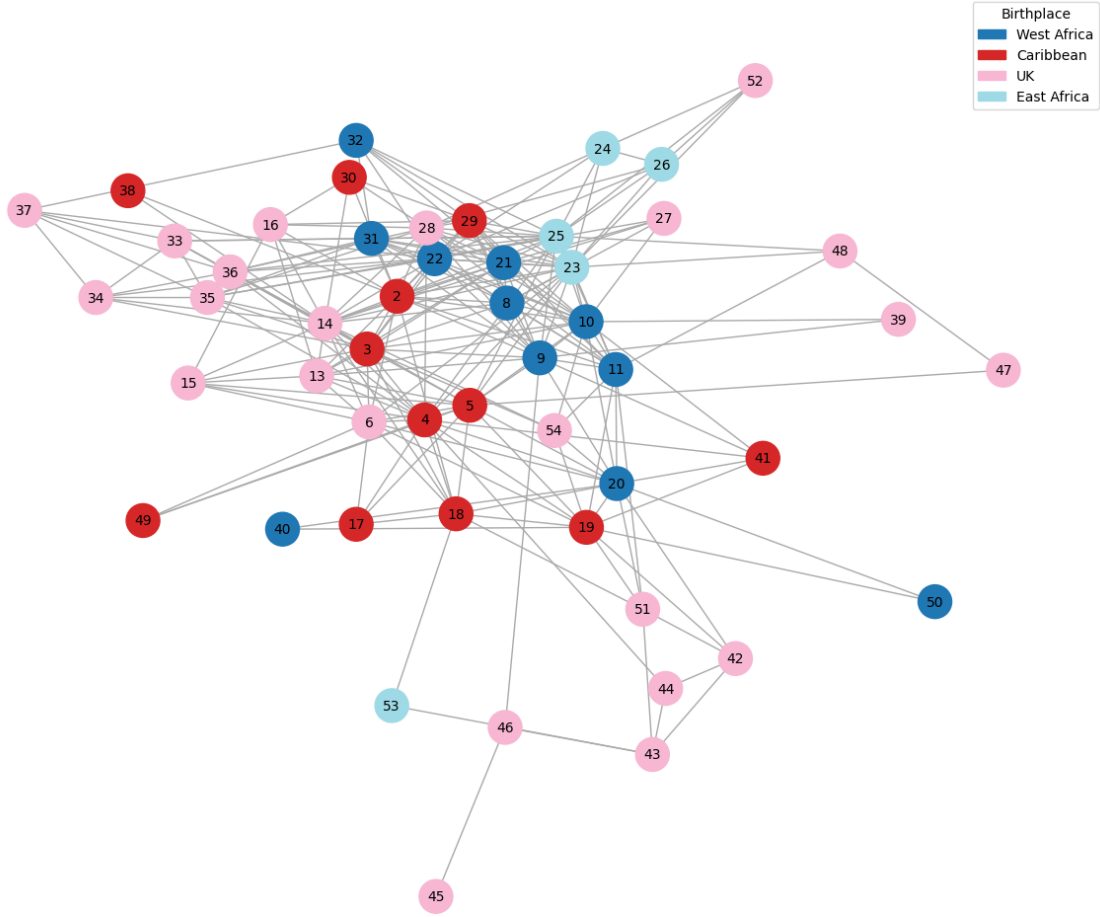


Figure 6: London Network sub-graph visualization without leaders.

5.2.4 Synthesis: Connection Between Data, Measures, and Properties

The **properties** are what we discovered: a "small-world" network (path length 2.05), highly cohesive (density 0.22), locally clustered (clustering 0.63), but undivided (modularity 0.26). It possesses a clear hierarchy (k-core 11) and defined social roles (Leaders 1, 7, 12), all of which combine to make it exceptionally **robust**.

5.2.5 Ethnicity analysis

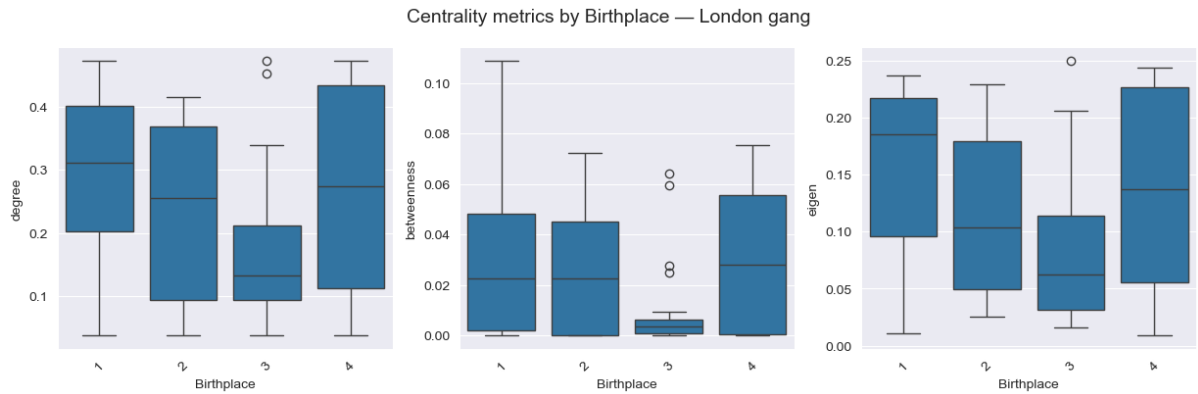
The analysis of the London gang network shows a **weak but positive tendency toward ethnic homophily**, with an assortativity coefficient of 0.113. This indicates that individuals display a mild preference for forming ties with others sharing the same *Birthplace*, although the overall network remains relatively integrated.

The **mixing matrix** confirms this observation: diagonal values (particularly for groups 1 and 3) are slightly higher, indicating intra-group cohesion, while off-diagonal entries remain substantial. This balance highlights the presence of numerous **cross-ethnic connections** within the gang's structure.

Table 9: Mixing matrix (proportion of connections between Birthplace groups) – London gang

| | 1 | 2 | 3 | 4 |
|---|-------|-------|-------|-------|
| 1 | 0.111 | 0.060 | 0.076 | 0.041 |
| 2 | 0.060 | 0.073 | 0.076 | 0.025 |
| 3 | 0.076 | 0.076 | 0.146 | 0.043 |
| 4 | 0.041 | 0.025 | 0.043 | 0.025 |

When analyzing **centrality measures**, groups 1 and 4 emerge as the most central and structurally influential, with the highest mean degrees (0.286 and 0.267 respectively) and eigenvector centralities (0.152 and 0.135). Group 2 follows closely, while group 3 (despite being the largest) exhibits the lowest centrality values, suggesting a more peripheral or clustered role. This pattern implies that influence and connectivity are distributed across multiple ethnic groups, rather than concentrated in a single one.



The **community analysis** identified four main communities, each with different degrees of ethnic diversity. Community 0 displays a relatively balanced composition (1: 38%, 2: 10%, 3: 29%, 4: 24%), while community 1 is dominated by groups 2 and 3. Communities 2 and 3 are less diverse, with community 2 almost entirely composed of group 3 members. The mean Shannon diversity index ($H = 0.886$) indicates moderate internal diversity, slightly lower than in the Italian network.

Table 10: Community composition by Birthplace – London gang

| Community | 1 | 2 | 3 | 4 |
|-----------|------|------|------|------|
| 0 | 0.38 | 0.10 | 0.29 | 0.24 |
| 1 | 0.20 | 0.35 | 0.40 | 0.05 |
| 2 | 0.00 | 0.14 | 0.86 | 0.00 |
| 3 | 0.00 | 0.33 | 0.67 | 0.00 |

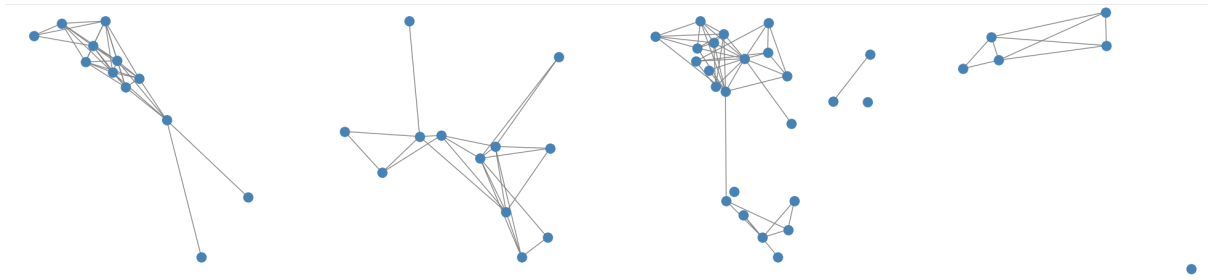
Furthermore, 64.44% of all connections occur between individuals of different *Birthplace* categories, demonstrating a high degree of cross-ethnic integration. As in the Italian case, national origin does not appear to be a key organizing principle in the network's structure.

Subgraph analysis by *Birthplace* provides additional insight:

- Groups **1** and **4** exhibit the highest internal density (0.530 and 0.533 respectively) and clustering, indicating strong intra-group cohesion.
- Group **3**, while the largest (24 nodes), has a lower internal density (0.167), suggesting looser internal connectivity and a more outward orientation.
- Group **2** shows intermediate density (0.348) and clustering, forming connections both internally and across groups.

Table 11: Subgraph-level statistics by Birthplace – London gang

| Birthplace | Nodes | Edges | Density | Clustering |
|------------|-------|-------|---------|------------|
| 1 | 12 | 35 | 0.530 | 0.658 |
| 2 | 12 | 23 | 0.348 | 0.636 |
| 3 | 24 | 46 | 0.167 | 0.506 |
| 4 | 6 | 8 | 0.533 | 0.722 |



In summary, the London gang network displays slightly lower diversity but comparable integration when compared to the Italian case. While certain groups form cohesive internal clusters, the overall structure is characterized by extensive cross-ethnic linkage and distributed influence across national backgrounds. Ethnicity, therefore, plays only a minor role in shaping the gang's internal connectivity patterns.

5.3 Comparison

General structural metrics

| Centrality Metric | Italian Network | London Network |
|--------------------------------|-----------------|----------------|
| Density | 0.0516 | |
| Average Degree | 3.0430 | |
| Diameter | 6 | |
| Average Path Length | 3.012 | |
| Average Clustering Coefficient | 0.4347 | |
| Number of communities | 5 | |
| Modularity Score | 0.5561 | |

Centrality Metrics

| Centrality Metric | Italian Network | | London Network | |
|------------------------|-----------------|--------|----------------|-------|
| | Top 3 nodes | Value | Top 3 nodes | Value |
| Degree Centrality | 19 | 0.3182 | | |
| | 63 | 0.2879 | | |
| | 18 | 0.2727 | | |
| Betweenness Centrality | 19 | 0.5558 | | |
| | 63 | 0.3633 | | |
| | 18 | 0.2881 | | |
| Closeness Centrality | 19 | 0.5397 | | |
| | 63 | 0.4597 | | |
| | 18 | 0.4563 | | |
| Eigenvector Centrality | 19 | 0.4394 | | |
| | 18 | 0.3784 | | |
| | 63 | 0.3110 | | |

Roles and Vulnerability

| | | Italian Network | London Network |
|---------------|------------------------------|---|----------------|
| Leaders | | 19, 63, 18, 4 | |
| Brokers | | 19, 63, 18, 47 | |
| Peripheral | | 2, 3, 5, 17, 22, 26, 33, 34, 36, 37, 38, 45, 46, 48, 49, 51, 53, 55, 56, 60, 61, 66, 67 | |
| Main core | k | 3 | |
| | Number of nodes | 20 (out of 67) | |
| | Nodes in the core | 4, 8, 11, 12, 13, 15, 18, 19, 21, 24, 31, 32, 39, 41, 44, 47, 58, 59, 63, 64 | |
| | Nodes | 24 (64 remaining) | |
| After Removal | Number of components | 18 | |
| | Average shortest path length | 3.1341 | |
| | density | 0.1014 | |
| | Node reduction | 64.18% | |
| | Increase in average path | 4.05% | |

6 Conclusion

Qualitative analysis of the quantitative findings of the study.

7 Critique

Do you think your work solves the problem presented above? To which extent (completely, what parts)? Why? What could you have done differently to answer your research problems (e.g., gather data with additional information, build your model differently, apply alternative measures)?

References

- [1] Ucinet covert networks datasets: Italian gangs. <https://sites.google.com/site/ucinetsoftware/datasets/covert-networks/italian-gangs>. Accessed 2025-11-11.
- [2] Ucinet covert networks datasets: London gang. <https://sites.google.com/site/ucinetsoftware/datasets/covert-networks/london-gang>. Accessed 2025-11-11.
- [3] J. McGloin and S. Decker. Theories of gang behavior and public policy. In H. Barlow and S. Decker, editors, *Criminology and Public Policy: Putting Theory to Work*, pages 150–165. Temple University Press, Philadelphia, 2010.
- [4] S. Decker, C. Katz, and V. Webb. Understanding the black box of gang organisation: Implications for involvement in violent crime, drug sales, and violent victimization. *Crime & Delinquency*, 54:153–172, 2008.
- [5] J. McIllwain. Organized crime: A social network approach. *Crime, Law and Social Change*, 32:301–323, 1999.
- [6] M. Klein and C. Maxson. *Street Gang Patterns and Policies*. Oxford University Press, New York, 2006.
- [7] A. D. Smith. *The Ethnic Origins of Nations*. Blackwell, Oxford, 1987.
- [8] C. Adamson. Defensive localism in white and black: A comparative history of European American and African-American youth gangs. *Ethnic and Racial Studies*, 23:272–298, 2000.
- [9] A. Alonso. Racialized identities and the formation of black gangs in Los Angeles. *Urban Geography*, 25:658–674, 2004.
- [10] A. Freng and F. Esbensen. Race and gang affiliation: An examination of multiple marginality. *Justice Quarterly*, 24:600–628, 2007.
- [11] D. Pyrooz, A. Fox, and S. Decker. Racial and ethnic heterogeneity, economic disadvantage, and gangs: A macro-level study of gang membership in urban America. *Justice Quarterly*, 27:867–892, 2010.
- [12] D. Starbuck, J. Howell, and D. Lindquist. Hybrid and other modern gangs, 2001.
- [13] Malcolm Klein and Lyle Crawford. Groups, gangs, and cohesiveness. *Journal of Research in Crime and Delinquency*, 30:63–75, 1967.
- [14] James F. Short and Fred L. Strodbeck. *Group Processes and Gang Delinquency*. University of Chicago Press, Chicago, IL, 1965.
- [15] Mark Fleisher. Fieldwork research and social network analysis: Different methods creating complementary perspectives. *Journal of Contemporary Criminal Justice*, 21:120–134, 2005.
- [16] Carlo Morselli. *Inside Criminal Networks*. Springer, New York, 2009.

- [17] Stanley Wasserman and Katherine Faust. *Social Network Analysis: Methods and Applications*. Cambridge University Press, Cambridge, 1994.
- [18] Aric A. Hagberg, Daniel A. Schult, and Pieter J. Swart. Exploring network structure, dynamics, and function using networkx. In Gaël Varoquaux, Travis Vaught, and Jarrod Millman, editors, *Proceedings of the 7th Python in Science Conference*, pages 11 – 15, Pasadena, CA USA, 2008.