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 [3]) and *London Gang Network* (54 nodes, 315 edges) (available online at [4]) 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 [11]. As a result, the connection between gang activity and gang structure has remained largely unexplored, described as a "black box" [5]. 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" [12].

Klein and Maxson [10] 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 [16] — play a central role in gang dynamics. Research has consistently shown that ethnicity is a key factor in both gang formation and membership [1] [2] [7] . While ethnic diversity has been recognized as influential in broader community contexts [14], its specific relationship to gang activity and internal organization remains largely unexplored [17].
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 [9] [15]. Technological advancements have significantly expanded its potential [6] [13]. 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 [18]. 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

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 [4]; the Italian dataset was obtained from the same repository [3]. Both include person–person adjacency matrices and accompanying node-attribute tables, and are fully digitised, requiring no manual transcription.

3.1 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`.

3.2 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

4.1 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.

4.2 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 files in [4] [3].
- **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.

Overall, while the datasets provide only partial depictions of real-world social structures, the analytical process applied to them is transparent, stable, and highly reproducible.

5 Measures and Results

For both the Italian and London gangs, we study the same network metrics and compare them. We start by analyzing **general structural metrics**, which quantify the overall organization and efficiency of the network:

General Structural Metrics

- **Density (D)**: measures how many ties exist in the network compared to the maximum possible number.

$$D = \frac{2m}{n(n-1)} \tag{1}$$

  where m is the number of existing edges and n the number of nodes. High density indicates strong cohesion and low vulnerability to node removal.

- **Average degree ($\langle k \rangle$)**: the mean number of connections per node.

$$\langle k \rangle = \frac{1}{n}\sum_{i=1}^{n} k_i = \frac{2m}{n} \tag{2}$$

  It represents member activity and overall engagement.

- **Network diameter (D_{max})** and **average path length (ℓ)**: respectively, the maximum and the mean of the shortest path distances between all node pairs.

$$\ell = \frac{1}{n(n-1)}\sum_{i \neq j} d_{ij} \tag{3}$$

  They describe the efficiency of information or order transmission within the network.

- **Clustering coefficient (C)**: expresses the probability that two neighbors of a node are connected, revealing local cohesion or closed "cells". For node i:

$$C_i = \frac{2t_i}{k_i(k_i-1)} \tag{4}$$

  where t_i is the number of triangles passing through i. The overall clustering coefficient is the average of C_i across all nodes:

$$C = \frac{1}{n}\sum_{i=1}^{n} C_i \tag{5}$$

- **Modularity (Q)**: quantifies the presence of well-defined communities:

$$Q = \frac{1}{2m}\sum_{i,j}\left(A_{ij} - \frac{k_i k_j}{2m}\right)\delta(c_i, c_j) \tag{6}$$

  where $\delta(c_i, c_j) = 1$ if nodes i and j belong to the same community. High Q indicates strong internal divisions or cliques.

Centrality Metrics

We also compute **centrality metrics**, which identify key actors based on different notions of importance:

- **Degree centrality** ($C_D(i)$): number of direct ties a node has.

$$C_D(i) = k_i \tag{7}$$

  It reveals the most active or visible members.

- **Betweenness centrality** ($C_B(i)$): measures how often a node lies on the shortest paths between others.

$$C_B(i) = \sum_{s \neq i \neq d} \frac{\sigma_{sd}(i)}{\sigma_{sd}} \tag{8}$$

  where σ_{sd} is the number of shortest paths between s and d, and $\sigma_{sd}(i)$ those passing through i. It highlights brokers or gatekeepers.

- **Closeness centrality** ($C_C(i)$): reciprocal of the mean distance from a node to all others.

$$C_C(i) = \frac{n-1}{\sum_{j \neq i} d_{ij}} \tag{9}$$

  Nodes with high C_C can quickly disseminate information.

- **Eigenvector centrality** ($C_E(i)$): assigns importance based on being connected to other important nodes.

$$C_E(i) = \frac{1}{\kappa} \sum_{j=1}^{n} A_{ij} C_E(j) \tag{10}$$

  where κ is the largest eigenvalue of the adjacency matrix A. It identifies leaders recognized by other influential members.

Roles and Network Hierarchy

Based on these metrics, we classify network roles as follows:

- **Leaders**: nodes in the top 5% for both degree and eigenvector centrality, broadly connected and influential.
- **Brokers**: nodes in the top 5% for betweenness, key intermediaries between subgroups.
- **Peripherals**: nodes in the bottom 5% for degree, few connections and limited influence.

We further analyze the k**-core decomposition** to inspect hierarchical structure and power concentration. A k-core is a maximal subgraph where each node has degree $\geq k$. The *main core* corresponds to the highest k with non-empty core.
We also assess **network robustness** by comparing metrics (density, average path length, k-core) before and after removing central nodes, in order to study the structural impact of node removal.

Attribute-based Network Analysis

Finally, to explore how ethnic background shapes internal organization, we perform **attribute-based network analyses** based on the *Birthplace* attribute.

- **Assortativity coefficient** (r) and **mixing matrix**: measure homophily by birthplace.

$$r = \frac{\sum_i e_{ii} - \sum_i a_i b_i}{1 - \sum_i a_i b_i} \tag{11}$$

  where e_{ij} is the fraction of edges linking group i to j, and a_i, b_i are the row and column sums of the mixing matrix.

- **Centrality by group**: comparing average degree, betweenness, and eigenvector centrality across ethnic categories reveals whether certain groups occupy more central positions.

- **Community composition and diversity**: community detection via modularity maximization; ethnic heterogeneity measured with the **Shannon Diversity Index**:

$$H = -\sum_i p_i \log(p_i) \tag{12}$$

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

- **Inter-group connectivity**: proportion of edges linking individuals of different birthplace categories — indicating cross-ethnic integration.

- **Subgraph analysis by birthplace**: evaluation of intra-group cohesion through internal density and clustering coefficient.

Note: Since the Italian network is disconnected, metrics such as diameter and average path length are computed on the largest connected component, including cases where central nodes are removed.

5.1 Italian gang

5.1.1 Structural analysis

Using the Library Networkx [8] 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.

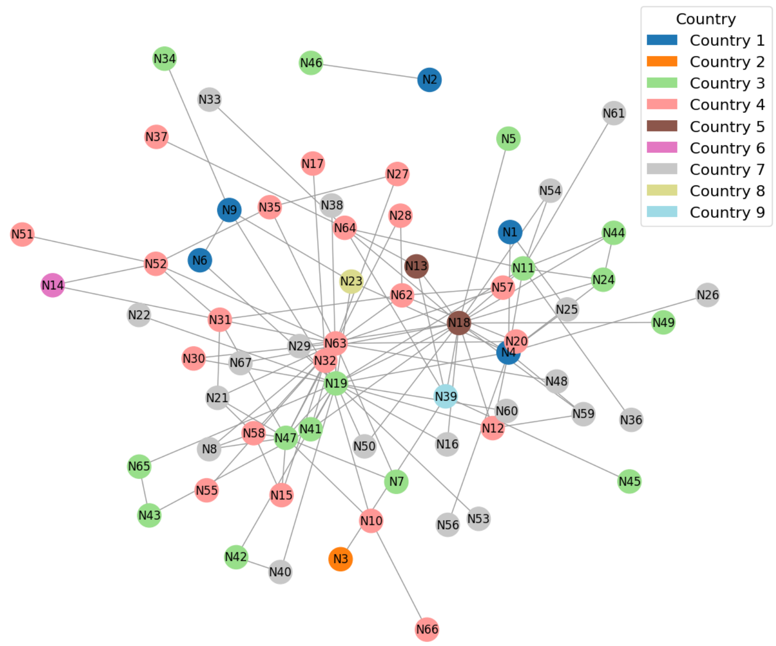


Figure 1: Italian Network graph visualization. Each node is colored according to its country label

5.1.2 Macro-level Cohesion and Structure

As the italian network is not connected and diameter and average path length are not defined for not connected networks, we instead calculate these metrics for the largest connected component, which has 65 nodes (out of 67).

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 (largest connected component)	3.012	Any two members can reach each other in about 3 "hops" on average.
Diameter (largest connected component)	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	0.0152	0.0000
3	0.0152	0.0000	0.2443	0.0132
5	0.0152	0.0000	0.3119	0.0556
17	0.0152	0.0000	0.2463	0.0141
22	0.0152	0.0000	0.3487	0.0645
26	0.0152	0.0000	0.2941	0.0384
33	0.0152	0.0000	0.2473	0.0179
34	0.0152	0.0000	0.2619	0.0140
36	0.0152	0.0000	0.2941	0.0349
37	0.0152	0.0000	0.2473	0.0179
38	0.0152	0.0000	0.3487	0.0645
45	0.0152	0.0000	0.2443	0.0132
46	0.0152	0.0000	0.0152	0.0000
48	0.0152	0.0000	0.3134	0.0457
49	0.0152	0.0000	0.3119	0.0556
51	0.0152	0.0000	0.2434	0.0117
53	0.0152	0.0000	0.3487	0.0645
55	0.0152	0.0000	0.3134	0.0457
56	0.0152	0.0000	0.2941	0.0384
60	0.0152	0.0000	0.3487	0.0645
61	0.0152	0.0000	0.2941	0.0349
66	0.0152	0.0000	0.2324	0.0069
67	0.0152	0.0000	0.3119	0.0556

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]`.

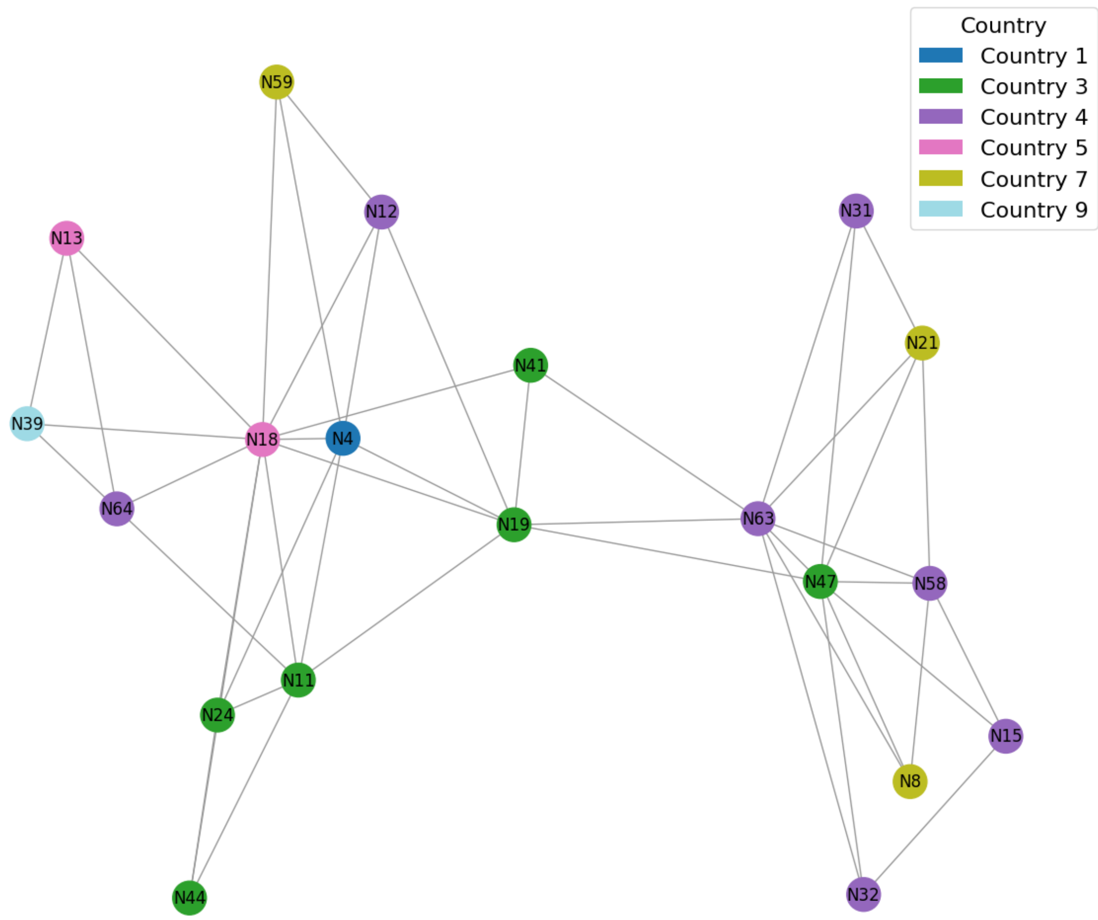


Figure 2: Core of Italian Network graph visualization.

Vulnerability Simulation Removing the three identified leaders (19, 63, 18) had a clear impact on the network structure. After their removal, the network broke into 18 components, and the largest connected part dropped from 65 to 24 nodes, showing a strong loss of cohesion. The density decreased from 0.0516 to **0.0288**, meaning fewer remaining connections overall. The average shortest path length increased from 3.0120 to **3.1341** (a **4.05%** rise), indicating slightly longer communication chains among the surviving nodes. Overall, the simulation shows that leader nodes play an important structural role, and removing them makes the network significantly more fragmented and less efficient.

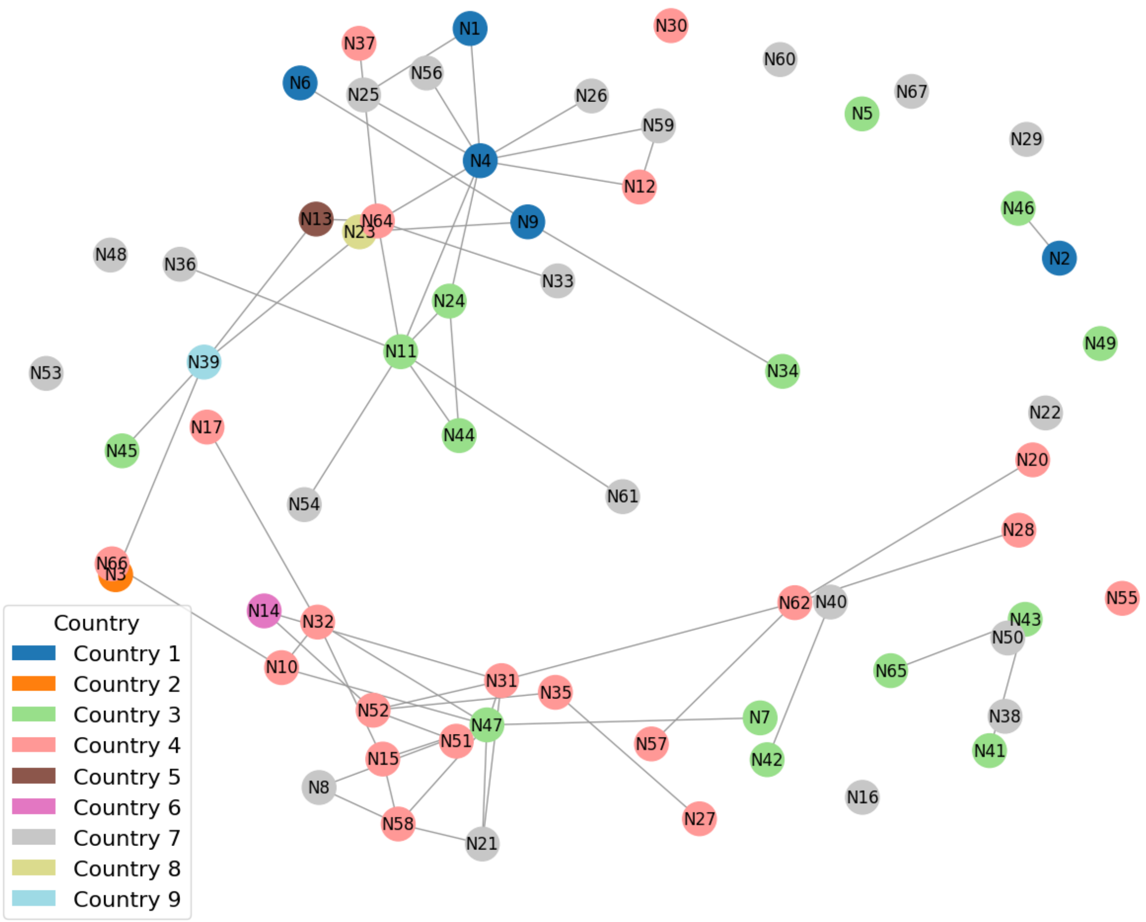


Figure 3: Italian Network sub-graph visualization without leaders.

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

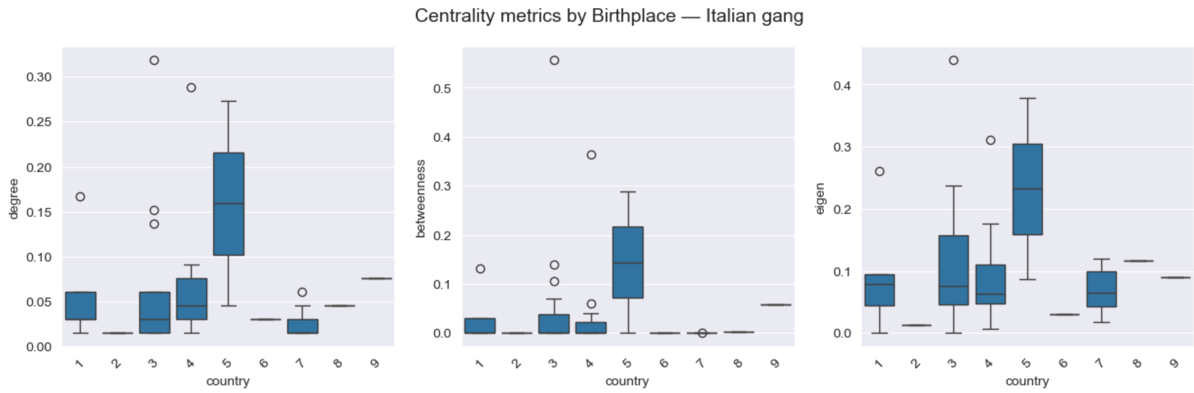
The analysis of the Italian gang network shows a **moderate tendency toward ethnic homophily**, with an assortativity coefficient of 0.150. This value suggests that individuals show a slight preference for connecting with others sharing the same *Birthplace*, though overall the network remains relatively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

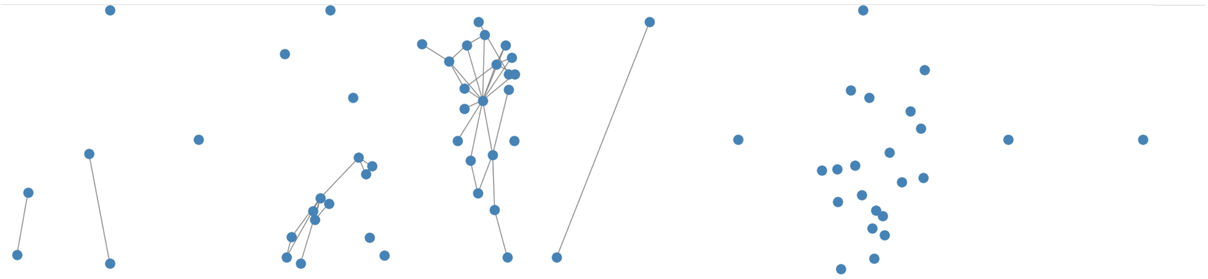
Subgraph analysis by *Birthplace* reveals additional information:

- 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

5.2.1 Structural analysis

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.

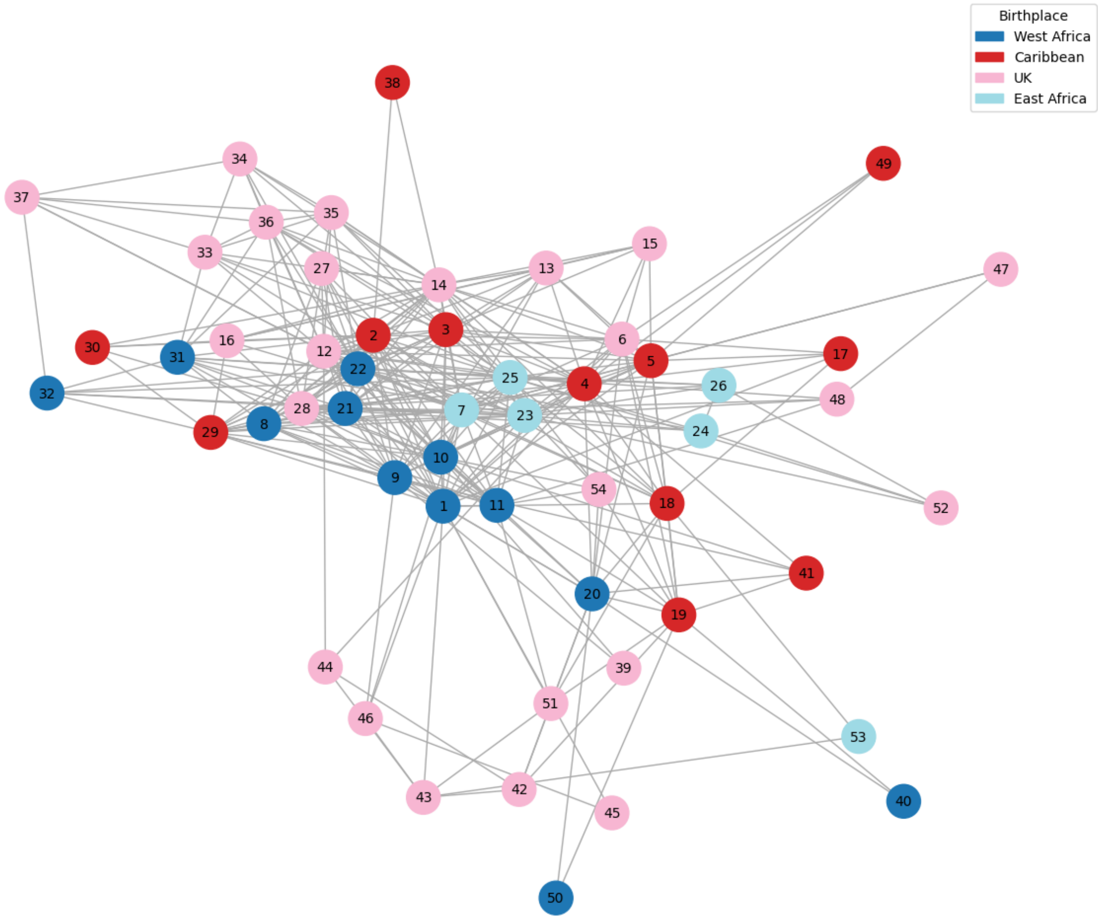


Figure 4: London Network graph visualization. Each node is colored according to its birthplace label

5.2.2 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.

5.2.3 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
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)

5.2.4 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 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]`.

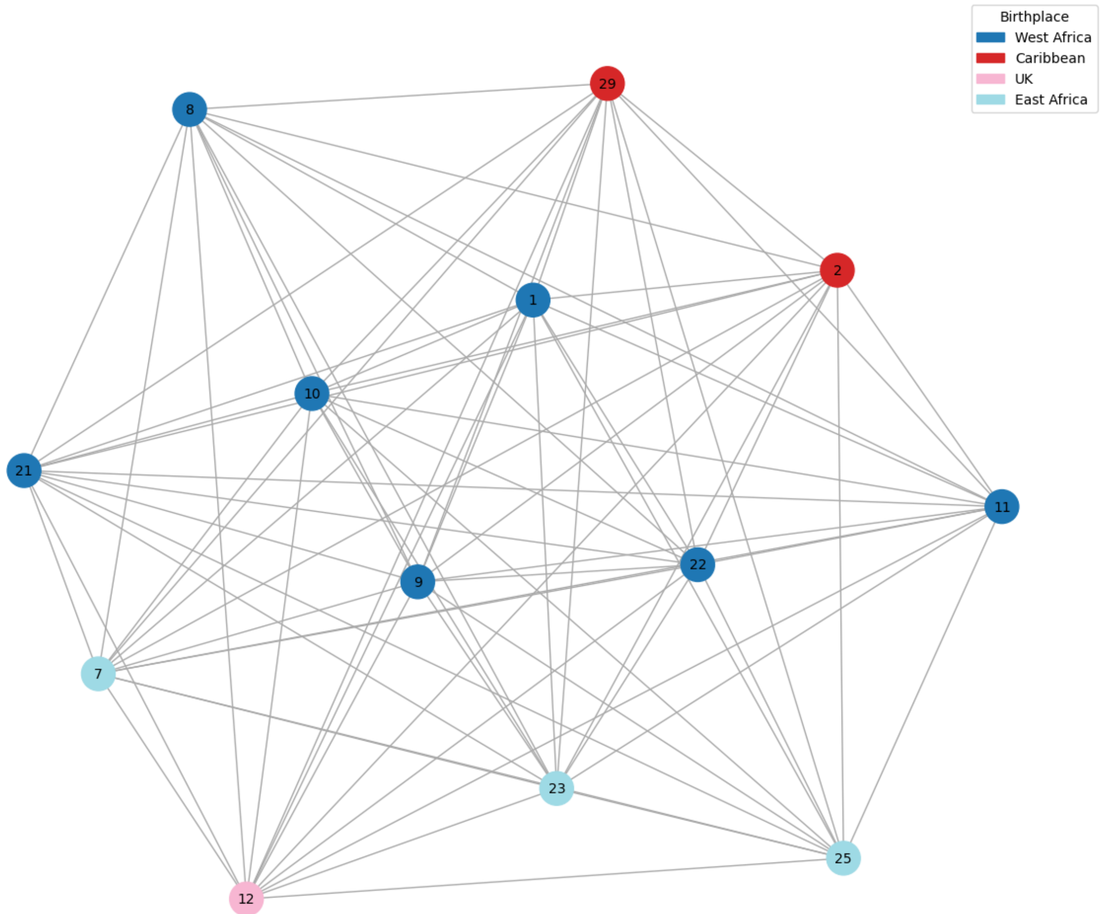


Figure 5: Core of London Network graph visualization.

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.

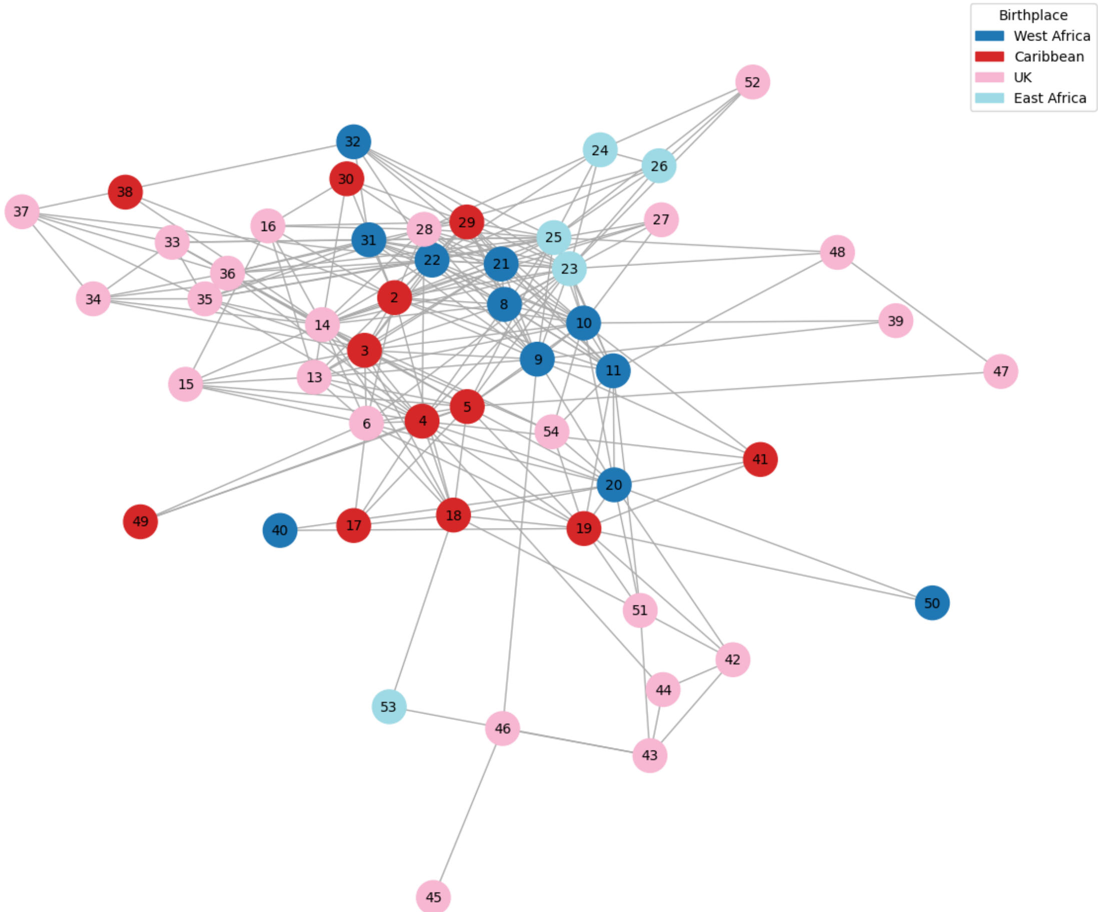


Figure 6: London Network sub-graph visualization without leaders.

5.2.5 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**.

5.2.6 Ethnicity analysis

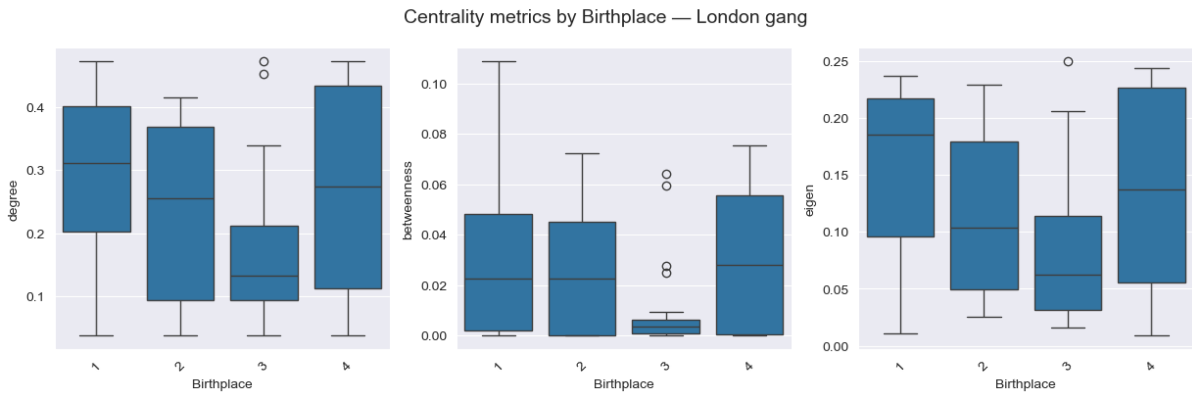
The analysis of the London gang network shows a **weak but positive tendency toward ethnic homophily**, with an assortativity coefficient of 0.113. This indicates that individuals display a mild preference for forming ties with others sharing the same *Birthplace*, although the overall network remains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

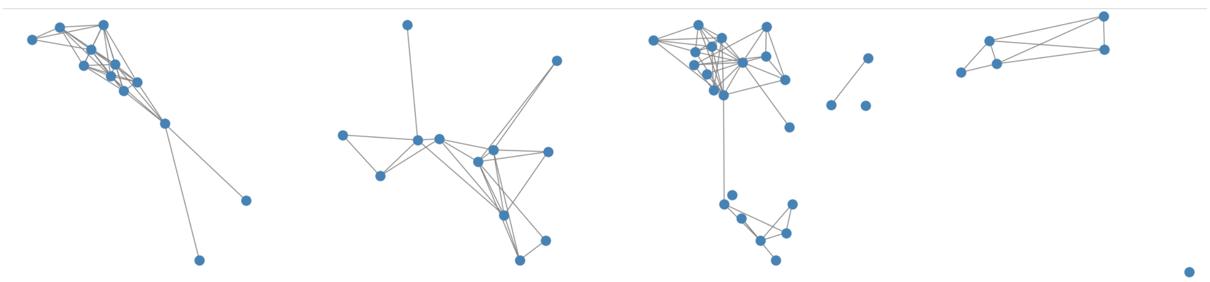
Furthermore, 64.44% of all connections occur between individuals of different *Birthplace* categories, demonstrating a high degree of cross-ethnic integration. As in the Italian case, national origin does not appear to be a key organizing principle in the network's structure.

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	
After Removal	Nodes	24 (64 remaining)	
	Number of components	18	
	Average shortest path length	3.1341	
	density	0.1014	
	Node reduction	64.18%	
	Increase in average path	4.05%	

5.4 Comparison

This section presents a comparative analysis of the two covert criminal networks, in which we examine their structural metrics in order to understand the distinct organizational structures of each group.

5.4.1 Network Topology and Cohesion

The structural properties of the two networks reveal markedly different organizational patterns: the Italian Network is considerably sparser than the London one, a difference evident in both its lower **density** (Italian: 0.0516, London: 0.2201) and its smaller **average degree** (Italian: 3.0430, London: 11.67). This suggests that interpersonal ties in the Italian gang are more selectively-distributed, and that interactions are far less intense, while the London gang functions as a highly integrated, tightly-knit community.

Consistently, the Italian network exhibits a longer **average path length** (Italian: 3.012, London: 2.05) and a larger **diameter** (Italian: 6, London: 4), indicating a lower network efficiency. On the other hand, the London Gang constitutes a more structurally efficient environment in which information and communication can propagate more quickly, requiring, on average, only two intermediaries.

Local cohesion reinforces this distinction: the **clustering coefficient** is higher in the the London network (London: 0.6331, Italian: 0.4347), indicating a network that is richer in transitive relationships. The most telling difference in the topology of the networks lies in the community structure. The Italian network displays higher **modularity** (Italian: 0.5561, London: 0.2665), which suggests it is organized into distinct sub-groups or cells. These modules are typically internally cohesive but are only sparsely connected to each other. On the other hand, the the London gang's low modularity, combined with its high density, indicates a lack of significant sub-divisions. It operates as a single, large, cohesive core group rather than a confederation of smaller cells.

5.4.2 Centrality Measures: Hierarchy, Leadership and Brokerage

The analysis of centrality measures of the two networks highlights a clear hierarchical difference. In particular, a clear and steep hierarchy can be observed in the Italian network. The three identified leaders (nodes 19, 63, 18) possess high scores across all centrality measures, while the identified peripheral members are 30 times less central. This aligns with the network's high modularity: these leaders function as **brokers**, bridging the structural holes between the otherwise separated cells.

In the London Gang, the hierarchy appears flatter, as it shows a more balanced distribution of centrality. In particular, two core actors (nodes 1 and 7) occupy both leadership and brokerage positions, but the overall gap between such central nodes and the peripheral ones is smaller (approximately one order of magnitude lower).

Furthermore, the identified leaders have high **degree** and **closeness**, but relatively low **betweenness**, which shows they act as "hubs" in the center of a single, dense core, rather than proper brokers connecting disparate groups.

5.4.3 K-Core and Robustness

The K-Core analysis provides an additional insight into the robustness of these two organizations. The Italian network culminates in a modest 3-core composed of 20 nodes, reinforcing its sparse nature.

As for robustness, removing the identified leaders caused a noticeable increase in average path length and a slight reduction in density, indicating the central role of these individuals in maintaining short communication distances within the core.

On the contrary, the London network exhibits a much more cohesive, robust structure, featuring a 11-core composed of 13 nodes, which suggests a group with strong internal redundancy and resilience: even after removing key leaders, the network density remains fairily large and the increase in average path length is modest.

5.4.4 Ethnicity Analysis

This section examines how members' ethnicity influences the relational structure of each group, with the aim of discovering to which extent sociocultural factors contribute to the internal functioning of each organization.

While both networks operate in distinct geographical contexts, they share a foundational characteristic: a high degree of integration across ethnic lies. More specifically, the tendency towards ethnic homophily, as measured by the assortativity coefficient, is positive, but remains weak in both cases. The Italian network exhibits a slightly higher coefficient (0.150) compared to the London network (0.113), yet the proximity and the magnitude of the two values indicate that while members may have a slight preference for forming ties with individuals of the same origin, ethnicity harmony is not a dominant structural driver in either case.

This observation is strongly supported by the proportion of cross-ethnic connections. The two networks are remarkably similar in this regard, with (64.91%) of ties in the Italian gang and (64.44%) of ties in the London gang occurring between individuals of different ethnicity groups.

This surprising similarity highlights a common underlying dynamic: both organizations are fundamentally integrated and heavily rely on cross-national cooperation, rather then acting as ethnically segmented groups. The mixing matrices for both networks corroborate these observations: each matrix contains high diagonal values for specific ethnic groups, yet both exhibit non-negligible cross-diagonal values.

Despite this high-level similarity, the two networks exhibit a key difference in their internal community structure, highlighted by the community analysis. In particular, the Italian network, which features a higher Shannon Diversity Index of H = 1.174, demonstrates an overall higher level of internal heterogeneity, meaning that most communities integrate members from multiple birthplaces, with only one (Community 4) being clearly dominated by two ethnic groups.

Conversely, the London Network exhibits lower community diversity, with a Shannon diversity index of H = 0.886, showing that several communities (mainly communities 1, 2 and 3), show strong ethnic concentrations. This pattern indicates that, while cross-ethnic ties remain common at the macroscopic level, the members of this network tend to cluster into ethnically homogeneous groups at the community level.

Concerning the distribution of influence of the different ethnic groups, it can be observed that the Italian network exhibits a relatively concentrated influence structure, with ethnicity group 5 clearly emerging as the most central and influential. In contrast, the London network exhibits a more distributed influence structure, where centrality is not monopolized by a single ethnic group. Instead, it is shared primarily by ethnic groups 1 and 4, with group 2 also playing an important role.

6 Conclusion

Our comparative investigation ultimately reveals two distinct models of covert criminal organizations which, however, share underlying structural patterns and properties.

First, concerning the networks' topologies and their global structural and organizational properties, our analysis demonstrated that the Italian gang operates as a more decentralized, modular organization, characterized by a higher modularity and a steeper hierarchy. On the other hand, the London gang functions as a denser, more cohesive centralized core, with a lower modularity and a flatter leadership structure.

Second, it emerged from this study that, contrary to similar studies [] which deem ethnicity as a primary driver the formation of ties between gangs members, both networks demonstrated a fundamental functional reliance on cross-ethnic ties, evidenced by low ethnic homophily and solid rates of cross-ethnic ties.

An important distinction lies in the community structure of the two networks: the Italian network maintains high ethnic diversity within the communities, but presents a single, dominant ethnic group, while the London network, despite its overall integration, internally clusters into more ethnically homogeneous communities.

In conclusion, this analysis highlights that ethnicity is not a dominant force able to shape tie formation or dictate interactions in the analyzed networks, but rather interacts complexly with the other network characteristics to shape patterns of cooperation, influence and community organization.

7 Critique

This study effectively provided meaningful insights into both key objectives, successfully outlining the core topological and structural features of both networks, and demonstrating that ethnicity is not a unique, monolithic driver for ties between members in the analyzed network.

However, these findings are inevitably constrained by the quality of the data itself. As discussed in the Validity section, the datasets represent simplified, static an somehow incomplete models of the real organizations, so the reconstructed topologies are necessarily approximations, especially considering that covert networks are only partially observable. Specifically, the choice to treat both networks as unweighted allowed for an easier comparison, but limited the fidelity of the analysis, since it flattened the distinction between weak interactions and strong ties. In addition, the two datasets differ in the richness and granularity of nodes metadata: the Italian one only provides a country-of-origin variable, while the London dataset contains numerous social and behavioral metadata, which again have been neglected to maintain comparability between networks, in spite of fidelity.

Another limitation arises from the static nature of the model: criminal networks evolve over time, yet the analysis was carried out on a single snapshot for each of the organizations. A temporal analysis could have offered a deeper understanding of the investigated phenomena, if the data had been available.

In addition, the data limitations observed here also point to a broader issue that extends beyond the data quality of the specific instances that have been examined: ethnic homophily in covert criminal networks is a complex phenomenon, whose analysis cannot be fully addressed through isolated network instances. Thoroughly addressing this goal requires a broader effort on a much larger and more detailed empirical basis, which incorporates more complex and attribute-richer datasets.

Despite the discussed limitations, the analysis still offers meaningful insights into the structural characteristics and dynamics of the examined networks, contributing to a clearer understanding on the organizational and interaction patterns of these groups, and providing a useful foundation to future research.

References

[1] Christopher Adamson. “Defensive localism in white and black: A comparative history of European-American and African-American youth gangs”. In: *Gangs*. Routledge, 2017, pp. 143–169.

[2] Alejandro A Alonso. “Racialized identitites and the formation of black gangs in Los Angeles”. In: *Urban Geography* 25.7 (2004), pp. 658–674.

[3] *Analytic Technologies - Italian Gangs — sites.google.com*. `https://sites.google.com/site/ucinetsoftware/datasets/covert-networks/italian-gangs`. 2016.

[4] *Analytic Technologies - London Gang — sites.google.com*. `https://sites.google.com/site/ucinetsoftware/datasets/covert-networks/london-gang`. 2016.

[5] Scott H Decker, Charles M Katz, and Vincent J Webb. “Understanding the black box of gang organization: Implications for involvement in violent crime, drug sales, and violent victimization”. In: *Crime & delinquency* 54.1 (2008), pp. 153–172.

[6] Mark S Fleisher. “Fieldwork research and social network analysis: Different methods creating complementary perspectives”. In: *Journal of Contemporary Criminal Justice* 21.2 (2005), pp. 120–134.

[7] Adrienne Freng and Finn-Aage Esbensen. “Race and gang affiliation: An examination of multiple marginality”. In: *Justice Quarterly* 24.4 (2007), pp. 600–628.

[8] Aric A. Hagberg, Daniel A. Schult, and Pieter J. Swart. “Exploring Network Structure, Dynamics, and Function using NetworkX”. In: *Proceedings of the 7th Python in Science Conference*. Ed. by Gaël Varoquaux, Travis Vaught, and Jarrod Millman. Pasadena, CA USA, 2008, pp. 11–15.

[9] Malcolm W Klein and Lois Y Crawford. “Groups, Gangs, and Cohesiveness”. In: *Journal of Research in Crime and Delinquency* 4.1 (1967), pp. 63–75.

[10] Malcolm W Klein and Cheryl L Maxson. *Street gang patterns and policies*. Oxford University Press, 2006.

[11] Jean McGloin and Scott H Decker. “Theories of gang behavior and public policy”. In: *Criminology and public policy: Putting theory to work* (2010), pp. 150–165.

[12] Jeffrey Scott McIllwain. “Organized crime: A social network approach”. In: *Crime, law and social change* 32.4 (1999), pp. 301–323.

[13] Carlo Morselli. *Inside criminal networks*. Vol. 8. Springer, 2009.

[14] David C Pyrooz, Andrew M Fox, and Scott H Decker. “Racial and ethnic heterogeneity, economic disadvantage, and gangs: A macro-level study of gang membership in urban America”. In: *Justice Quarterly* 27.6 (2010), pp. 867–892.

[15] James F Short and Fred L Strodtbeck. “Group process and gang delinquency”. In: (1965).

[16] Anthony D Smith. *The ethnic origins of nations. New York, NY: B*. 1987.

[17] David Starbuck, James C Howell, and Donna J Lindquist. *Hybrid and other modern gangs*. Office of Juvenile Justice and Delinquency Prevention Washington, DC, 2001.

[18] Stanley Wasserman. “Social network analysis: Methods and applications”. In: *The Press Syndicate of the University of Cambridge* (1994).