Network Analysis of Musical Collaborations on Spotify: The Italian Music Scene

Federico Borci, Computer science, 0001130010
Federica Santisi, Computer science, 0001186853
Giorgia Pirelli, Computer science, 0001176116

January 9th, 2026

1 Introduction

The music industry has undergone radical transformations in the last two decades, shifting from a traditional model based on physical record sales to a digital ecosystem dominated by streaming platforms. Spotify, launched in 2008, has become the leading global music streaming service, with over 500 million active users and a catalog exceeding 100 million tracks.
In this new landscape, musical collaborations, commonly known as "featurings," have assumed a central role. Whereas in the past collaborations were relatively rare events often limited to special projects, today they represent a fundamental strategy for artists at every level of popularity. Collaborations enable artists to reach new audiences, experiment with different musical genres, increase their visibility on streaming platforms, and forge strategic connections within the music industry.
Social Network Analysis (SNA) offers powerful methodological tools for studying these relational dynamics. By representing artists as nodes and their collaborations as edges, it is possible to construct a network that captures the complexity of interactions in the contemporary musical landscape.
This study focuses on the analysis of collaborations among artists on Spotify, with a particular emphasis on the Italian music scene as the primary case study.

Research Questions

Through a network perspective, this work aims to answer fundamental questions about the functioning of the Italian musical ecosystem and its positioning in the international context:

1. **What are the fundamental structural characteristics of the Italian musical collaboration network on Spotify?** Global network metrics: number of nodes and edges, connected components, average and maximum degree, density, diameter, average shortest path length, clustering coefficient, and transitivity.

2. **How does the Italian scene position itself relative to major European and non-European countries?** Cross-country comparison using network size, number of collaborations, average degree, density, clustering coefficient, connected components, and node–edge linear regression.

3. **Who are the most central and influential artists, and what structural role do they occupy?** Degree, eigenvector, closeness, and betweenness centrality.

4. **How are connections distributed within the network?** Degree distribution analysis (linear and logarithmic scale).

5. **What are the prevalent collaboration patterns? Do artists tend to collaborate with peers similar in terms of connectivity, popularity, or musical genre?** Assortativity coefficients by degree, followers (popularity), and genre; intra- and inter-genre collaboration counts.

6. **Do well-defined communities exist in the network? How do they relate to musical genres?** Community detection (Louvain, Edge Betweenness), modularity, and genre purity.

7. **What collaborative strategies do emerging artists adopt compared to established artists?** Artist classification by popularity and collaborations; collaboration matrices between classes and centrality differences.

2 Datasets

The initial dataset used for this analysis was downloaded from Spotify Artist Feature & Collaboration Network [4]. After careful inspection, it was considered a reliable source, as it is derived from publicly available data provided by the Spotify API and already pre-processed to represent collaboration relationships among artists. The resulting dataset is modeled as an undirected, attribute-enriched network, where node-level metadata enable the investigation of social, cultural, and structural mechanisms.
The dataset consists of two main files:

- **nodes.csv** – contains the nodes of the graph, where each node represents an artist. The columns include:
  - `id`: unique identifier of the artist.
  - `name`: name of the artist.
  - `followers`: number of followers of the artist on Spotify.
  - `popularity`: popularity index (0–100), computed by Spotify based on recent streams and overall visibility.
  - `genres`: list of genres associated with the artist.
  - `chart_hits`: List showing the number of Spotify chart hits in different countries.
- **edges.csv** – contains the edges of the graph, representing a collaboration between two artists. The columns include:
  - `source`: ID of the collaborating artist.
  - `target`: ID of the artist being collaborated with.

The initial objective was to enrich the graph with additional artist-level information, specifically **nationality** and **dominant musical genre**, in order to enable more in-depth social and cultural analyses of the network.

2.1 Artist Nationality Enrichment

Nationality is not directly observable in the original dataset and is therefore treated as an inferred node attribute. The enrichment strategy was designed to maximize coverage while minimizing systematic misclassification bias. To associate a nationality with each artist, two complementary strategies were adopted:

1. **Inference based on musical genre.** In the first approach, nationality was inferred by analyzing the associated musical genres. For instance, an artist labeled with the genre `italian hip hop` was classified as *Italian*. This method allowed the automatic assignment of nationality to a substantial subset of artists; however, it was not applicable in all cases, as many genres do not contain explicit geographical references.

2. **Completion using an external dataset (MusicBrainz).** For artists whose nationality could not be inferred in the first step, data from the `MusicBrainz Dump (mbdump)` [5] were integrated. A direct matching based solely on artist names posed significant challenges due to the presence of homonyms with different nationalities. To mitigate this issue, the integration was performed exclusively on artists that remained unclassified after the first inference step.

This hybrid procedure increased the coverage of nationality information and enabled a more accurate subsequent analysis, particularly when comparing artistic communities across different countries.

2.2 Addition of Musical Genre

To associate one or more musical genres with artists in the dataset, a multi-level procedure was designed with the goal of maximizing coverage while preserving semantic consistency. Each stage operates only on artists that remained unclassified after the previous step.

- **Direct mapping of Spotify genres.** Spotify-specific genres were first normalized and mapped to a limited set of musical macro-categories (e.g., *Pop*, *Rock*, *Hip Hop / Rap*, *Electronic / Dance*) using lexical and keyword-based rules.

- **AI-assisted completion.** Genres that could not be mapped automatically were classified through an AI-assisted process, which assigned them to the predefined macro-categories based on semantic similarity.

- **Inference via artistic collaborations.** For artists still lacking a genre, the collaboration network was exploited by assigning the most frequent genres among direct collaborators. When necessary, this inference was extended using a Breadth-First Search (BFS) up to three levels.

- **Popularity-based inference.** Remaining unclassified artists were analyzed using quantitative indicators such as follower count and popularity, inferring the most likely genres based on patterns observed in the dataset.

- **Global fallback assignment.** In the few remaining cases, a fallback strategy based on the most common genres in the dataset was applied.

This multi-level strategy enabled the creation of a dataset that is complete and consistent from a musical genre perspective, minimizing arbitrary assignments and integrally exploiting semantic, structural, and quantitative information. The final result constitutes a solid foundation for the subsequent network and artistic community analyses.

3 Validity and Reliability

The dataset provides a realistic representation of musical collaborations on Spotify, as it is derived from Spotify API data and models collaborations as reciprocal relationships between artists. This abstraction is suitable for capturing structural properties of the contemporary music collaboration ecosystem. This modeling choice primarily supports the analysis of structural properties, while abstracting away semantic and contractual aspects of collaborations. Limitations arise from platform-specific biases and data incompleteness, since not all collaborations or dimensions of artistic influence are observable on Spotify. Additionally, inferred attributes such as nationality and genre may introduce minor approximations, especially for artists with hybrid identities. These effects are mitigated by conservative, multi-stage enrichment procedures designed to reduce systematic bias. Reliability is ensured through the use of public data sources and deterministic preprocessing pipelines. All enrichment steps follow clearly defined and ordered rules, allowing the analysis to be replicated with the same inputs.
The only partially non-deterministic component is the AI-assisted genre classification; however, it is applied to a limited subset and its outputs are fixed before analysis, preserving reproducibility of the results. This choice slightly affects construct validity but does not compromise reproducibility, as classifications are fixed prior to network analysis.

4 Measures and Results

4.1 General Analysis of the Italian Musical Collaboration Network

In order to outline the key structural differences and to obtain an overview of the topology and internal dynamics of the Italian musical collaboration network, a series of general metrics were calculated, as reported in Table 1.

Table 1: Structural characteristics of the Italian musical collaboration network

Parameter	**Value**
Total nodes (artists)	1,656
Total edges (collaborations)	4,307
Connected components	16
Maximum node degree	114
Average node degree	5.20
Density	0.00314
Diameter	10
Average shortest path length	4.14
Average clustering coefficient	0.119
Transitivity	0.128

The sixteen connected components suggest the existence of isolated groups, corresponding to niches or artistic communities with limited contact with the rest of the national ecosystem. On average, each artist is connected to about five colleagues in the network. However, the degree distribution is highly heterogeneous: the presence of a node with degree 114 reveals a central *hub* of major importance, while 54.6% of artists have a degree of 1, thus collaborating with only one other artist. The extremely low density (approximately 0.31% of possible connections are realised) confirms the **sparse** nature of the network. The network diameter is 10, indicating that any two artists can be connected through at most 10 intermediate collaborations.
Despite the low density, the network presents a modest average path length (4.14), indicating that artists are connected through few intermediate steps.
The values of the average clustering coefficient (0.119) and transitivity (0.128) are moderate and close to each other. This indicates a measurable, though not dominant, tendency towards **triadic closure**: two collaborators of the same artist have approximately a 12% probability of having collaborated with each other in turn. This local cohesion fosters the formation of cohesive artistic circles and partially clustered communities, contributing to the stability of collaborative relationships and the sharing of artistic practices within subgroups, while maintaining sufficient openness to allow connections between different communities.

4.1.1 Comparative Analysis with Major European Countries

To situate the Italian results within a broader continental context, the analysis was extended to major European countries. This approach enables a comparative evaluation of collaborative dynamics. Table 2 presents the comparison between Italy and the top 5 European countries by network size.

Table 2: Comparison with the top 5 European countries by network size

Metric	Italy	United Kingdom (1st)	Germany (2nd)	France (4th)	Netherlands (5th)
Total nodes	1,656	3,290	2,706	1,643	1,420
Total edges	4,307	7,532	5,929	4,754	5,143
Average degree	5.20	4.58	4.38	5.79	7.24
Connected components	16	70	32	27	12
Average clustering	0.119	0.062	0.120	0.113	0.151
Density	0.00314	0.001392	0.001618	0.003525	0.005107

The analysis reveals significant structural differences among countries. The **Netherlands** exhibit the highest **average degree** in Europe (7.24) despite ranking only fifth in terms of number of artists. Their **average clustering coefficient** (0.151) is the second highest, and the percentage of hub artists (17.8%) is the maximum in Europe. The **United Kingdom**, despite its larger size (3,290 artists), presents the lowest **average clustering coefficient** (0.062) and the highest fragmentation (70 **connected components**), reflecting a vast but segmented market. It also shows a significantly lower network density (0.001392) compared to that observed for Italy (0.003143). Italy's position within the European context is intermediate. With 1,656 artists, the Italian network ranks third in size, but its **average degree** (5.20) is lower only than that of the Netherlands and France, indicating an intense propensity for collaboration per artist. Its **density** (0.00314) is significantly higher than that of much larger networks such as the United Kingdom and Germany, suggesting a more compact ecosystem. The moderate fragmentation (16 components) and the **average clustering** (0.119) place Italy in a balanced position between the high cohesion of Poland and Greece and the low cohesion of the United Kingdom.

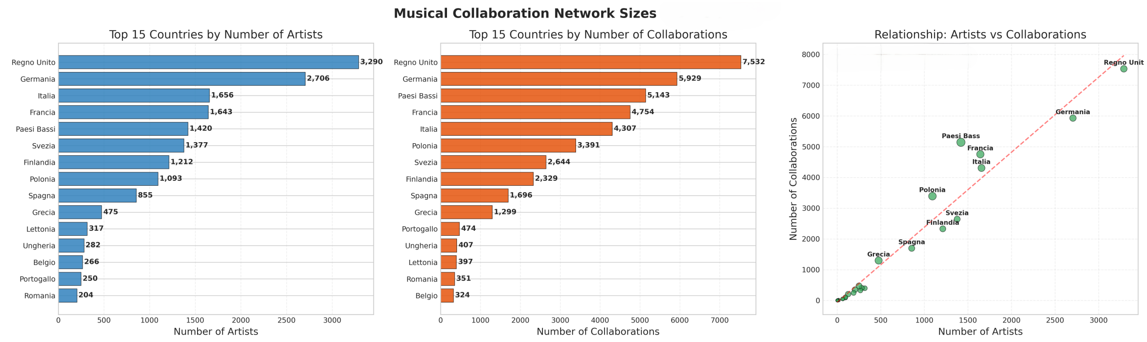


Figure 1: Comparative multidimensional analysis of musical collaboration networks at the European level.

Figure 1 presents a comparative multidimensional analysis of musical collaboration networks at the European level. The analysis synthesises three critical aspects of the structure of European networks: size, collaborative intensity, and the relationship between these variables. The histograms present the top 15 European countries for two distinct metrics: on the left, the number of artists; in the centre, the total number of collaborations. In the latter, it is observed that the **United Kingdom** (7,532 collaborations) and **Germany** (5,927 collaborations) rank at the top for the overall number of collaborations; this result is primarily attributable to their high number of nodes. The **Netherlands** (5,143 collaborations) represent the most relevant case: despite being fifth in number of artists, they rank third in collaboration volume, surpassing both Italy (4,307) and France (4,754). This indicates intense collaborative activity. **Poland** (3,391 collaborations) shows surprising intensity, surpassing countries with larger networks such as Sweden (2,644) and Finland (2,329).
The chart on the right explores the fundamental relationship between a network's size (number of artists) and its activity (number of collaborations). The relationship between the number of artists and the total number of collaborations was analysed using linear regression estimated by the least squares method, considering all European countries included in the study. Countries positioned **above the regression line** (Netherlands, Poland, Greece) represent hyper-collaborative ecosystems, where the volume of interactions systematically exceeds expectations given their dimensions. Countries **close to the line** (Italy, France, Germany, United Kingdom) follow an approximately linear relationship between collaborative activity and size.

4.1.2 Comparative Analysis with Major Non-European Countries

To situate the Italian musical ecosystem in a broader global context, a comparative analysis was conducted, extended to 49 non-European countries. Table 3 presents the comparison between Italy and the top 5 non-European countries by network size.

Table 3: Comparison with the top 5 non-European countries by network size

Metric	**Italy**	**United States (1st)**	**Brazil (2nd)**	**India (3rd)**	**Japan (4th)**	**Mexico (5th)**
Total nodes	1,656	6,217	1,859	999	890	773
Total edges	4,307	14,860	5,547	2,834	1,103	1,665
Average degree	5.20	4.78	5.97	5.67	2.48	4.31
Connected components	16	111	3	2	38	4
Average clustering	0.119	0.082	0.173	0.181	0.042	0.156
Density	0.00314	0.000769	0.003212	0.005685	0.002788	0.005580

The **United States** exhibit the largest network, with 6,217 artists and 14,860 collaborations, over 3.7 times larger than the Italian network. This scale is accompanied by very low density (0.000769) and a high number of connected components (111), indicating a highly diversified and fragmented music market. Despite its smaller size, Italy shows a higher average degree (5.20 vs. 4.78), suggesting more intense collaborative activity. **Brazil** and **India** emerge as large and highly cohesive musical systems, characterised by few connected components, and strong clustering.
In terms of collaborative intensity, Italy ranks among the most active countries globally, surpassing the United States, Japan, Mexico and South Korea in average degree. Several countries display hyper-collaborative structures: **Puerto Rico** represents the most extreme case, while **Brazil**, **India**, and **South Korea** combine relatively large networks with strong internal interconnectedness.
The analysis of network density reveals clear regional patterns. The analysed **African** countries (including Ghana and Algeria) exhibit very small but extremely dense networks, reflecting highly compact music scenes. **Latin American** countries show intermediate densities, still substantially higher than those observed in European and North American networks. Italy (0.00314) occupies an intermediate position: denser than the United States, comparable to Brazil, but less dense than India and Mexico, indicating a balanced trade-off between network size and cohesion.
Finally, the highest **clustering coefficients** are observed in **Puerto Rico**, **Ghana**, and the **Dominican Republic**, followed by **India** and **Brazil**, which combine large network size with strong local cohesion. Italy displays a moderate clustering value (0.119), higher than those of the United States and Japan but lower than the major emerging countries. This intermediate position suggests a musical ecosystem that effectively balances local cohesion and structural openness.

4.2 Nodes

4.2.1 Centrality Measures

To identify the structurally most important artists in the collaboration network, four centrality measures were calculated. The objective is to understand which artists occupy strategic positions and how these positions manifest themselves through different aspects of the network's structure.

4.2.1.1 Degree Centrality

Degree centrality quantifies the number of direct connections of a node, normalized by the maximum possible number of connections. In the analyzed musical network, this measure represents the number of distinct artists with whom a given artist has collaborated.
The implementation calculates both the absolute degree and the normalized degree centrality. The results reveal a strongly asymmetric distribution: the average degree centrality is 0.0031, while the maximum value reaches 0.0689. This distribution indicates that the majority of artists maintain a limited number of collaborations, while a small group of nodes concentrates a significantly high number of connections.

Table 4 presents the ten artists with the highest degree centrality:

Table 4: Top 3 artisti per Degree Centrality

Posizione	Artista	Degree Centrality
1	Guè	0.0689
2	Andrea Bocelli	0.0622
3	Clementino	0.0508

Guè emerges as the most connected artist, with 114 distinct collaborations, confirming his central role in the Italian music network. Andrea Bocelli ranks second with 103 collaborations, highlighting his ability to connect across different musical genres. The strong presence of hip hop artists among the top positions reflects the collaborative nature of this genre and its structural relevance within the network. Ennio Morricone's inclusion in the top ten introduces an element of stylistic and generational diversity.

4.2.1.2 Eigenvector Centrality

Eigenvector centrality assigns importance not only to the quantity of connections but also to their quality: an artist has high eigenvector centrality if they are connected to other artists who themselves occupy central positions in the network. The algorithm converges iteratively, assigning each node a score proportional to the sum of the scores of its adjacent nodes.
The results show a highly concentrated distribution, with an average value of 0.0089 and a maximum of 0.2573, indicating the presence of a small and cohesive core of highly influential artists.

Table 5: Top 3 artists by Eigenvector Centrality

Rank	Artist	Eigenvector Centrality
1	Guè	0.2573
2	Gemitaiz	0.2069
3	Emis Killa	0.1904

Table 5 presents the three artists with the highest eigenvector centrality. Guè clearly dominates the ranking, with a value approximately 0.2573 higher than the second-ranked artist, confirming his central role within the network. The ranking is exclusively composed of hip hop artists, revealing a genre-specific core that defines the structural center of the collaboration network. Artists such as Gemitaiz, Marracash, Jake La Furia, Don Joe, MadMan, and Lazza form a densely interconnected group. Notably, Andrea Bocelli is absent from this ranking despite his high degree centrality, suggesting that his collaborations are less embedded in the network's central core.

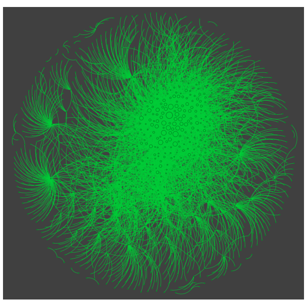


Figure 2: Gephi representation of Eigenvector

4.2.1.3 Closeness Centrality

Closeness centrality measures how close a node is to all other nodes in the network by computing the inverse of the average shortest-path distance. In the context of musical collaborations, an artist with high closeness centrality can reach any other artist through a small number of intermediaries, indicating a structurally advantageous position for communication and interaction.
The distribution shows an average value of 0.2336 and a maximum of 0.3677, suggesting that even the most central artists require only a few steps to access the entire network.

Table 6: Top 3 artists by Closeness Centrality

Rank	Artist	Closeness Centrality
1	Guè	0.3677
2	Clementino	0.3576
3	Gemitaiz	0.3537

Table 6 reports the three artists with the highest closeness centrality. Guè again occupies the top position, confirming his optimal placement within the network. In addition to the artists reported in the table, relatively high closeness centrality values are also observed for artists such as Clementino and J-AX, who appear more prominent in this metric than in eigenvector centrality. This pattern suggests a bridging role between different regions of the network rather than strong membership in its densest core. The presence of artists like Elisa and Rocco Hunt further highlights the ability of the network to support efficient connectivity across genres. Overall, closeness centrality identifies artists who facilitate rapid information flow and collaboration opportunities across the entire musical ecosystem.

4.2.1.4 Betweenness Centrality

Betweenness centrality measures how often a node lies on the shortest paths between other pairs of nodes. In a collaboration network, artists with high betweenness act as intermediaries connecting different clusters, even if they do not have many direct collaborations. For computational efficiency, the metric was approximated by sampling 1,000 nodes.
The distribution is highly skewed, with an average value of 0.0018 and a maximum of 0.1250, indicating that only a small number of artists exert substantial control over the flow of connections between different network communities. Andrea Bocelli clearly dominates this metric, highlighting his role as a structural bridge between classical, pop, and other musical genres. In addition to the artists reported in Table 7, relatively high betweenness values are also observed for artists such as Ennio Morricone, DJ Matrix, Jovanotti, and Cristina D'Avena, confirming that betweenness centrality captures a bridging role that is distinct from local or popularity-based measures. Clementino further stands out by combining high centrality across multiple metrics, effectively linking the hip hop core with other regions of the network.
Table 7 lists the three artists with the highest betweenness centrality.

4.2.2 Assortativity

Assortativity analysis examines connection tendencies based on specific node attributes. Three forms of assortativity were calculated to characterize collaboration patterns in the Italian musical network.

Table 7: Top 3 artists by Betweenness Centrality

Rank	Artist	Betweenness Centrality
1	Andrea Bocelli	0.1250
2	Clementino	0.0910
3	Guè	0.0642

4.2.2.1 Degree Assortativity

The degree assortativity coefficient of the network is **-0.1052**, indicating a slightly **disassortative** structure. This negative value shows that highly connected artists tend to collaborate with less connected artists, rather than forming tight clusters among themselves.
In the musical network, this pattern suggests that hubs do not form an isolated core but include less prolific artists in their collaborations. This may reflect several dynamics: established artists providing visibility to emerging talent, producers and featuring artists working across a wide spectrum of collaborators, or a generally open music scene where past collaborations do not limit future opportunities. Although moderate, the negative assortativity supports a structure that promotes mobility and access for less central artists, contrasting with rigid hierarchical networks where only the most successful figures collaborate among themselves.

4.2.2.2 Followers Assortativity

The followers assortativity coefficient is **0.0724**, a positive but near-zero value, indicating a very weak tendency towards assortative patterns based on popularity. This metric measures whether artists with a similar number of Spotify followers preferentially collaborate with each other.
The result suggests that popularity has limited influence on collaboration patterns. Highly followed artists do not exclusively collaborate with similarly popular artists, nor do less popular artists remain isolated. Categorizing edges into High-High, Low-Low, and High-Low collaborations using the median follower count shows a relatively balanced distribution, with both intra- and inter-popularity collaborations occurring.
This indicates a fluid Italian music scene, where fame does not strongly constrain partnerships. While the slightly positive coefficient hints at a weak homophilic tendency very popular artists collaborating somewhat more frequently with other popular artists it is not strong enough to create significant structural barriers within the network.

4.2.2.3 Genre Assortativity and Modularity

The musical genre assortativity coefficient is **0.4778**, a clearly positive value indicating a strong tendency for artists to collaborate within the same or related genres. The modularity is **Q = 0.2987**, moderate according to the interpretative scale, showing that while genre strongly influences collaborations, cross-genre connections still maintain global network cohesion.
Table 8 lists the ten most frequent genre pairs. Hip Hop / Rap dominates with 1829 intra-genre collaborations, followed by Pop with 819. The Hip Hop / Rap + Pop combination is the most frequent cross-genre connection (596 edges), acting as a structural bridge between the two major genres. Other notable cross-genre collaborations include Hip Hop / Rap + Indie and Elettronica / Dance + Pop, though less frequent. Smaller yet cohesive communities are found in Classica / Orchestrale, Elettronica / Dance, and Soundtrack / Film Score.

Table 8: Top 10 genre pairs in collaborations

Rank	Genre 1	N. collab.	Genre 2
1	Hip Hop / Rap	1829	(intra-genere)
2	Pop	819	(intra-genere)
3	Hip Hop / Rap	596	Pop
4	Hip Hop / Rap	104	Indie
5	Elettronica / Dance	85	(intra-genere)
6	Classica / Orchestrale	72	(intra-genere)
7	Elettronica / Dance	62	Pop
8	Indie	45	Pop
9	Soundtrack / Film Score	37	(intra-genere)
10	Elettronica / Dance	34	Hip Hop / Rap

Overall, the results reveal a pattern of genre specialization with selective cross-genre interactions. While artists tend to collaborate predominantly within their stylistic boundaries, established bridges, especially between hip hop and pop, ensure network connectivity. Hip Hop / Rap's high internal density confirms its central role and highly interconnected internal ecosystem.

4.3 Network

4.3.1 Community Detection

Community detection was applied to assess genre-based collaboration patterns. Two methods were considered, namely the Louvain algorithm and the Edge Betweenness approach. For each method, the number of detected communities and their genre purity were evaluated.

Louvain

The Louvain algorithm identified a total of 34 communities, revealing a relatively fragmented network structure. Several communities exhibit high genre homogeneity, particularly for *Hip Hop / Rap* and *Pop*, with purity values exceeding 0.6 and reaching 1.0 in smaller clusters. This pattern is visually reflected in the Louvain-based Gephi layout (Figure 3). Overall, this indicates that collaborations are not strictly constrained by genre boundaries, especially within larger communities.

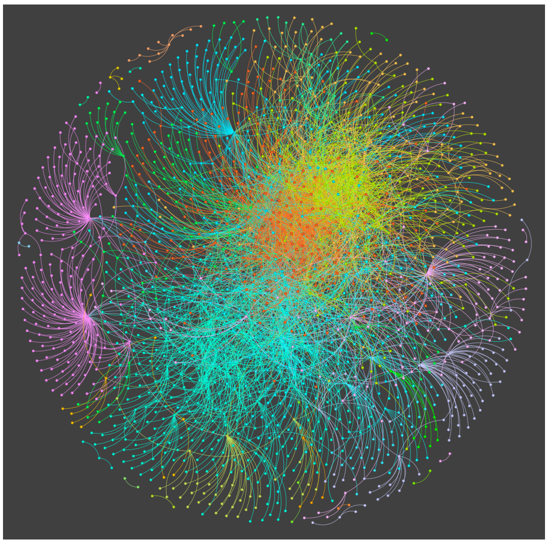


Figure 3: Community structure obtained by applying the Louvain algorithm in Gephi.

Louvain versus Genre Assortativity

The community detection results can be interpreted using genre assortativity and modularity measures. The genre assortativity coefficient ($r = 0.4778$) indicates a clear homophilic tendency, with artists more likely to collaborate within the same or closely related macro-genres.

The moderate modularity value ($Q = 0.2987$) suggests that genre does not induce a strong structural separation, due to frequent cross-genre collaborations, notably between *Hip Hop / Rap* and *Pop*.
Overall, the network displays a pattern of *selective mixing*: local genre-based homophily coexists with cross-genre ties, resulting in a structure that is both cohesive and interconnected.

Edge Betweenness

The Edge Betweenness (Girvan–Newman) algorithm produced 17 communities, resulting in a coarser partition than Louvain. Communities are generally less pure, with genre purity often below 0.5, highlighting the role of bridge nodes and cross-genre collaborations rather than clear genre separation.

4.3.2 Degree Distribution

The analysis of the degree distribution provides insight into the global structure of the artist collaboration network. The minimum degree of 1 reflects the presence of artists involved in a single collaboration, while the maximum degree of 114 highlights a small set of highly connected nodes acting as hubs. The average degree of 5.20 indicates an overall sparse network.
This heterogeneous connectivity pattern is visible in the Gephi visualization (Figure 4), where node size is proportional to degree, indicating a strongly right-skewed distribution. This structure is typical of complex networks, where a small number of hubs play a central role in maintaining global connectivity.

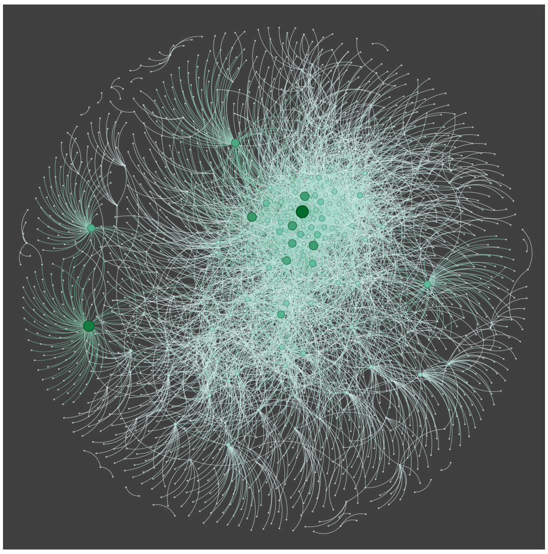
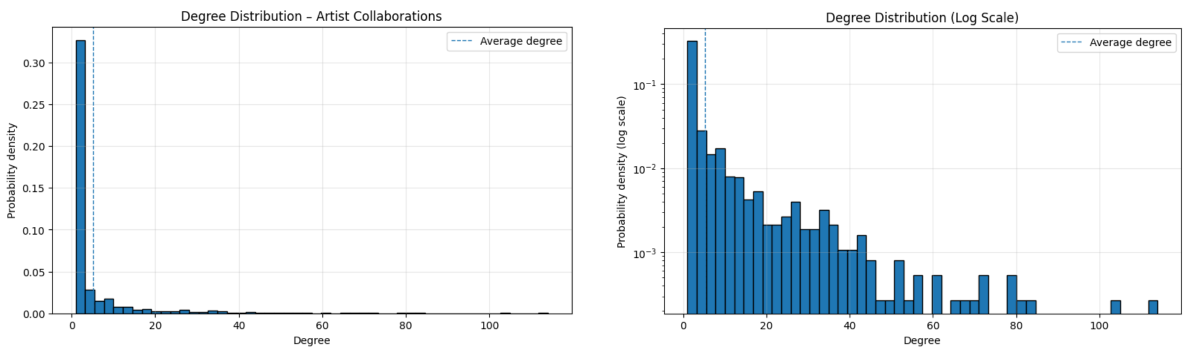


Figure 4: Artist collaboration network visualized in Gephi with node size proportional to degree.



(a) Degree distribution in linear scale.

(b) Degree distribution in logarithmic scale.

Figure 5: Degree distribution of the artist collaboration network.

5 Conclusion

This network analysis reveals that the Italian mu-
sical collaboration ecosystem on Spotify exhibits a complex and stratified structure. The network is characterized by **high sparsity** (density 0.00314) but efficient "small-world" connectivity (diameter 10, average path length 4.14). The degree distribution is highly heterogeneous: while the majority of artists (54.6%) have only one collaboration, a small core of central hubs notably **Guè**, **Andrea Bocelli**, and **Clementino** concentrates a very high number of connections, supporting global cohesion.
Centrality measures show an influence concentration within a **cohesive core dominated by Hip Hop/Rap** (evident in eigenvector centrality), while betweenness centrality highlights strategic bridge nodes (e.g., Bocelli, Morricone) connecting distinct stylistic communities such as classical, pop, and hip hop. The Italian network occupies an **intermediate range globally** in size (1,656 nodes, 6th among non-European countries) but stands out for its **above-average collaborative propensity** (average degree 5.20), positioning itself above the European trend in the node-edge regression.
Assortativity analysis indicates **strong genre homophily** (coefficient 0.4778), with collaborations predominantly intra-genre, especially in hip hop (1,829 edges) and pop (819 edges). However, significant inter-genre bridges particularly between Hip Hop/Rap and Pop (596 edges) ensure **permeability and global cohesion**, as confirmed by moderate modularity ($Q = 0.2987$). The structure is slightly disassortative by degree (-0.1052), suggesting that highly connected artists frequently collaborate with less central ones, fostering integration and mobility.
In summary, the Italian music scene on Spotify emerges as a **dynamic and balanced ecosystem**, characterized by a hyper-collaborative yet permeable core that combines local specialization with cross-genre openness, establishing itself as a relevant case study within the global digital music landscape.

6 Critique

Despite the analysis having produced significant and interpretable results, it is important to acknowledge the main intrinsic limitations of our methodological approach and the data used.

- **Partial and platform-limited data**: The dataset includes exclusively collaborations officially registered on Spotify, excluding those occurring on other platforms (YouTube, SoundCloud), in live contexts, or in unofficial forms. This may lead to an underestimation of network density, especially for underground genres or emerging artists operating outside mainstream channels.

- **Automatic attribute inference**: Nationality and musical genre were assigned through automated procedures. While this maximized coverage, it may have introduced systematic errors or excessive simplifications.

- **Static versus dynamic analysis**: The network was treated as a fixed snapshot in time. A longitudinal approach would have allowed studying how communities form, how artists' centrality changes in response to events (album release, chart entry), and how collaborative strategies evolve across different career stages.

- **Simplicity of the network model**: The graph is unimodal (only artists) and undirected, and does not distinguish between occasional collaborations and stable partnerships. A richer representation, perhaps weighted according to the number of shared tracks or enriched with temporal metadata, would have enabled more accurate analysis.

- **Interpretive limits of the structural approach**: Quantitative analysis identifies connection patterns; one can measure that an artist connects two musical communities, but it is not known whether this is a conscious creative choice, a personal relationship, or a commercial strategy. Without integrating structural data with qualitative sources, such as interviews, lyric analysis, reconstruction of production contexts, or market dynamics, conclusions about "collaborative strategies" or the "social role" of artists remain plausible yet not overly detailed hypotheses.

References

[1] *Network Analysis Project Repository*, GitHub, `https://github.com/edefbo1/Network_Analysis.git`

[2] Gephi Consortium, *Gephi: an open source graph visualization and analysis software*, Gephi.org, `https://gephi.org/`

[3] *NetworkX Documentation and Tutorial*, NetworkX.org, `https://networkx.org/documentation/stable/tutorial.html`

[4] J. Freyberg, *Spotify Artist Feature Collaboration Network*, Kaggle Dataset, `https://www.kaggle.com/datasets/jfreyberg/spotify-artist-feature-collaboration-network/`

[5] MusicBrainz Foundation, *MusicBrainz Database — Download and Documentation*, MusicBrainz.org, `https://musicbrainz.org/doc/MusicBrainz_Database/Download`