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

Federico , matricola
Federica Santisi, matricola
Giorgia Pirelli, matricola

December 2025

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 dataset is structured as a undirected graph and 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

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, thereby improving overall precision and preserving data consistency.

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.

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.

4 Measures and Results

4.1 General Analysis of the Italian Musical Collaboration Network

In order to outline the key structural differences and 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.

The **number of connected components** was measured to identify the presence of isolated subgroups within the national musical ecosystem. Both the **maximum degree** and the **average degree** of nodes were calculated to assess the network's global connectivity and to quantify the intensity of artists' collaborative activity. The network's **density** was calculated to measure its overall cohesion. The **diameter** and the **average shortest path length** provide an assessment of information flow efficiency and indicate the ease with which artists can reach each other. Finally, the **average clustering coefficient** and **transitivity** were calculated to measure the network's local cohesion and the tendency towards the formation of tightly-knit groups.

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

artistic niches with limited contact with the rest of the ecosystem. Each artist is connected, on average, to about five colleagues. However, the degree distribution is 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. The extremely low density (approximately 0.31% of possible connections) confirms the **sparse** nature of the network. Despite this, the network has a diameter of 10 and a modest average path length (4.14), indicating that artists are connected through few intermediate steps (a "small-world" structure).
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. This local cohesion fosters the formation of cohesive artistic circles, contributing to the stability of collaborative relationships within subgroups.

4.1.1 Comparative Analysis with Major European Countries

To situate the Italian results in a broader continental context, the analysis was extended to major European countries. Table 2 shows the top five European countries by number of artists.

Table 2: Comparison with major European countries

Metric	**Italy**	**France**	**Germany**	**United Kingdom**	**Netherlands**
Total nodes	1,656 (3rd)	1,643 (4th)	2,706 (2nd)	3,290 (1st)	1,420 (5th)
Total edges	4,307 (5th)	4,754 (4th)	5,929 (2nd)	7,532 (1st)	5,143 (3rd)
Average degree	5.20 (5th)	5.79 (2nd)	4.38 (7th)	4.58 (6th)	7.24 (1st)
Connected components	16 (4th)	27 (6th)	32 (7th)	70 (9th)	12 (3rd)
Average clustering	0.119 (4th)	0.113 (6th)	0.120 (3rd)	0.062 (9th)	0.151 (2nd)

The analysis reveals significant structural differences. The Netherlands present the highest average degree in Europe (7.24) despite ranking only fifth in number of artists, representing a hyper-collaborative and cohesive model (clustering 0.151). Countries like Poland and Greece represent models of highly cohesive and integrated networks, with high average degree and clustering and low fragmentation. At the opposite extreme, the United Kingdom, despite having the largest network (3,290 artists), presents the lowest clustering (0.062) and the highest fragmentation (70 connected components), reflecting a vast but segmented market. Its density (0.001392) is significantly lower than the Italian one (0.003143), indicating a more dispersed structure.
Italy positions itself in an intermediate range within the European context. Its average degree (5.20) is lower than that of the Netherlands and France, but higher than that of Germany and the United Kingdom. Its fragmentation (16 components) and its clustering (0.119) are moderate, showing an ecosystem balanced between local cohesion and openness.

Figure 1 synthesizes size, collaborative intensity, and the relationship between these variables. The central histogram shows how the Netherlands (5,143 collaborations), despite being fifth in number of artists, rank third in collaboration volume, surpassing Italy and France. The scatter plot on the right, with its regression line, highlights how countries like the Netherlands, Poland, and Greece position themselves above the line, showing a level of collaboration higher than expected given their size. Italy, together with France, Germany, and the United Kingdom, lies close to the line, following an approximately linear relationship between collaborative activity and size.

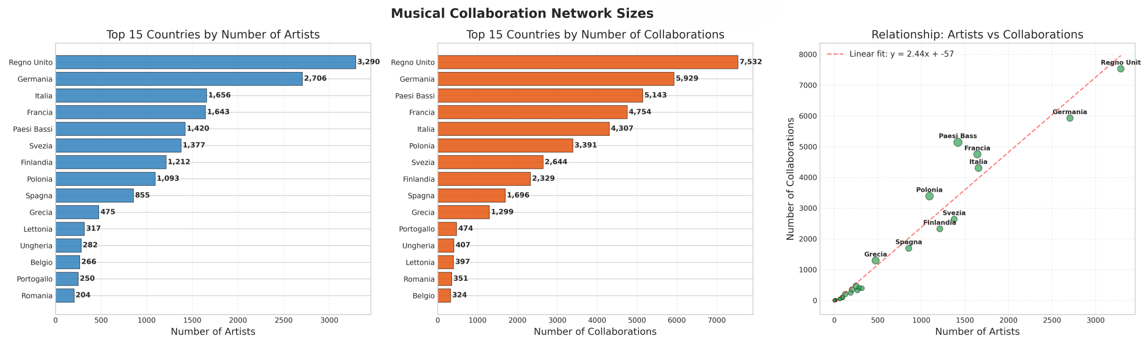


Figure 1: Comparative multidimensional analysis of musical collaboration networks at the European level.

4.1.2 Comparative Analysis with Major Non-European Countries

For a global context, the analysis was extended to 49 non-European countries. Table 3 compares Italy with 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 analysis reveals a marked dimensional asymmetry and different structural models. The United States present a network of exceptional size (6,217 artists) but with very low density (0.000769) and high fragmentation (111 components), reflecting a huge and diversified market; its average degree (4.78) is slightly lower than the Italian one. Brazil and India represent the major emerging ecosystems, characterized by high density and exceptional structural cohesion (only 3 and 2 components) and high clustering values (0.173 and 0.181). Italy ranks sixth among non-European countries by size. Compared to Japan (890 artists), Italy shows a much more cohesive structure (16 vs 38 components) and more intense collaborative activity (average degree 5.20 vs 2.48). With an average degree of 5.20, Italy positions itself among the countries with the highest collaborative activity globally, surpassing the United States, Germany, the United Kingdom, and South Korea.

Regarding density and fragmentation, the analyzed African countries represent the model of small, ultra-dense networks, with densities up to 40 times higher than the US one. Latin American networks show intermediate densities but significantly higher than European and North American ones. Italy, with a density of 0.00314, occupies an intermediate position: it is 4 times higher than that of the United States, similar to that of Brazil, about half that of India and Mexico, and higher than that of Japan.

With respect to clustering and local cohesion, the highest values are observed in Puerto Rico (0.314), Ghana (0.287), and the Dominican Republic (0.275). India (0.181) and Brazil (0.173) combine significant network size with high internal cohesion. Italy, with a value of 0.119, presents moderate but significant clustering in the global context: it is higher than that of

the United States (0.082) and Japan (0.042), but lower than that of India, Brazil, and Mexico (0.156). This indicates a network that balances local cohesion and openness. Very low clustering values in countries such as Canada (0.037) and Australia (0.035) suggest network structures less inclined towards triadic closure.
In conclusion, the Italian musical network configures itself as a medium-to-large sized ecosystem at the global level, characterized by an intense collaborative propensity (high average degree) and a structure that balances moderate local cohesion with contained fragmentation, positioning itself in an intermediate range.

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 reports the ten artists

Table 4: Top 10 artisti per Degree Centrality

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

with the highest degree centrality. 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 10 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 ten artists with the highest eigenvector centrality. Guè clearly dominates the ranking, with a value approximately 25

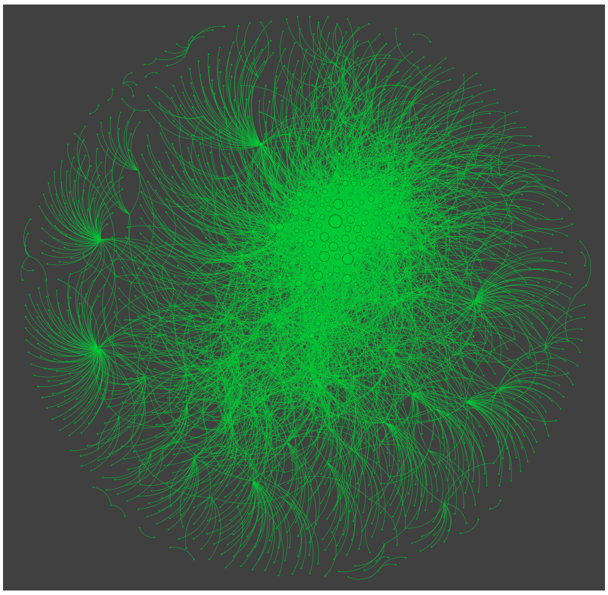


Figure 2: Gephi representation of Eigenvector

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 reports the ten artists with the highest closeness centrality. Guè again occupies the top position, confirming his optimal placement within the network. Artists such as Clementino and J-AX stand out more in this metric than in eigenvector centrality, indicating a bridging role between different network regions rather than membership in the densest core. The presence of Elisa and Rocco Hunt further highlights the ability of this network to support efficient

Table 6: Top 10 artists by Closeness Centrality

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

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 of 0.0018 and a maximum of 0.1250, indicating that only a few artists control the flow of connections between different network communities. Andrea Bocelli dominates this metric, highlighting his role as a structural bridge between classical, pop, and other genres. Artists such as Ennio Morricone, DJ Matrix, Jovanotti, and Cristina D'Avena further confirm that betweenness centrality captures a bridging role distinct from local centrality. Clementino combines high centrality in multiple metrics, linking the hip hop core with other parts of the network. Table 7 lists the ten artists with the highest betweenness centrality.

Table 7: Top 10 artists by Betweenness Centrality

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

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.

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: 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

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.

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 analysis was conducted to investigate whether artists tend to collaborate primarily with other artists belonging to the same musical macro-genre. To this end, two distinct approaches were applied, namely the Louvain algorithm and the Edge Betweenness (Girvan–Newman) method. For each approach, both the number of detected communities and their genre homogeneity were evaluated by measuring the purity of the dominant macro-genre within each community.

Louvain

The Louvain algorithm identified a total of 34 communities, revealing a relatively fragmented network structure. Several communities exhibit a high degree of genre homogeneity, particularly for the *Hip Hop / Rap* and *Pop* macro-genres, with purity values exceeding 0.6 and reaching 1.0 in smaller clusters. At the same time, many communities show a mixed composition, with multiple dominant macro-genres coexisting. This behavior is visually reflected in the community layout produced by the Louvain algorithm in Gephi (Figure 3), where dense, genre-centered clusters coexist with more diffuse, heterogeneous structures. Overall, this indicates that artist collaborations are not strictly constrained by genre boundaries, especially within larger communities where cross-genre interactions are more frequent.

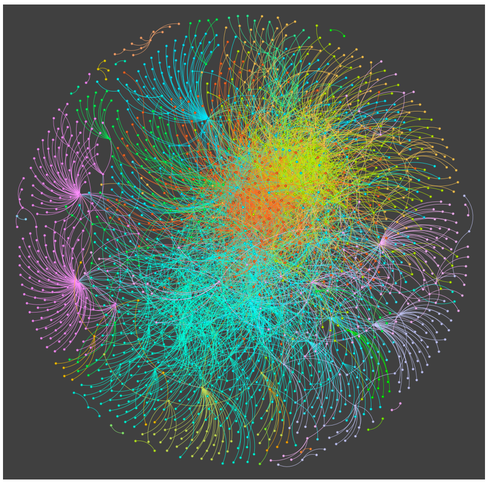


Figure 3: Community structure obtained by applying the Louvain algorithm in Gephi.

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. This behavior is reflected in the Louvain partition, which identifies high-purity communities, particularly for *Hip Hop / Rap* and *Pop*.
The moderate modularity value ($Q = 0.2987$) suggests that genre does not induce a strong structural separation of the network. Several communities exhibit a mixed genre composition, especially in the Edge Betweenness partition, where genre purity is often below 0.5. This pattern is explained by the presence of numerous cross-genre collaborations, notably between *Hip Hop / Rap* and *Pop*, which act as structural bridges and limit overall modularity.
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 partitioning of the network compared to the Louvain method. The identified communities are generally less pure, with genre purity values frequently below 0.5, particularly in larger clusters dominated by *Pop* and *Hip Hop / Rap*. This outcome indicates that the iterative removal of highly central edges tends to group together artists from different macro-genres, emphasizing the presence of bridge nodes and inter-genre collaborations rather than a clear separation based on musical genre.

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 clearly visible in the Gephi visualization (Figure 4), where node size is proportional to degree. A large number of small nodes coexist with a few prominent hubs, suggesting a strongly right-skewed distribution. This structure is typical of complex networks, in which highly connected nodes play a central role in maintaining global connectivity and facilitating interactions across different regions of the network, potentially spanning multiple musical genres.

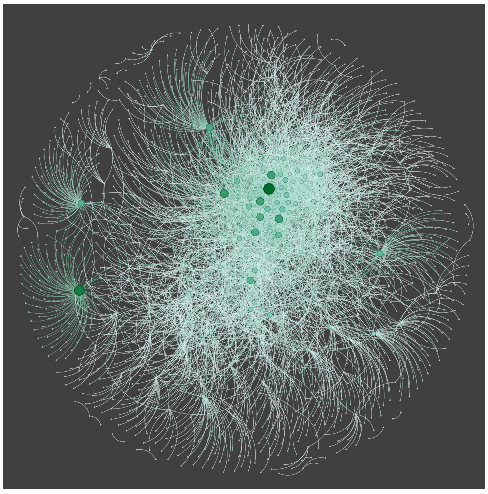
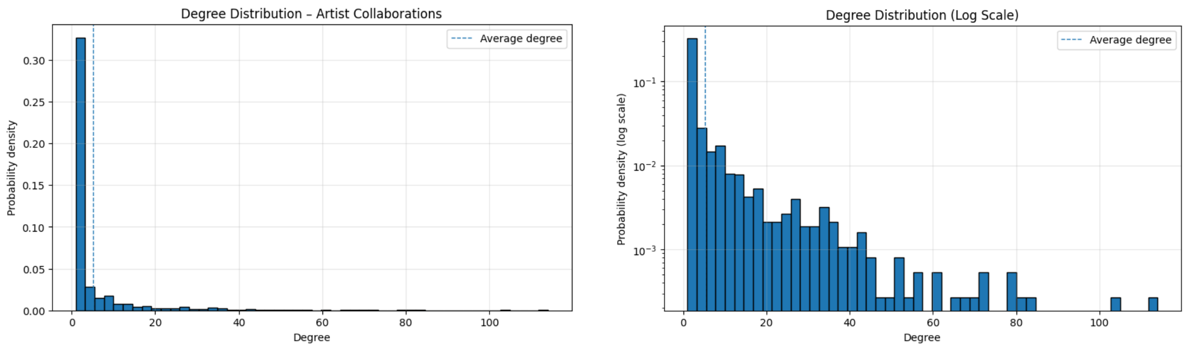


Figure 4: Artist collaboration network visualized in Gephi with node size proportional to degree.



(a) Degree distribution in linear scale. (b) Degree distribution in logarithmic scale.

Figure 5: Degree distribution of the artist collaboration network.

5 Conclusion

The investigation demonstrates that the graph of Italian musical collaborations on Spotify exhibits structural properties typical of real-world complex networks: high sparsity, heterogeneous degree distribution with a long tail, and the presence of central hubs. The quantitative analysis of centrality metrics reveals a significant concentration of relational power within a cohesive core dominated by the Hip Hop/Rap genre, while betweenness measures identify strategic bridge nodes connecting distinct stylistic communities. Assortativity and modularity coefficients indicate strong homophily based on genre, yet with sufficient permeability to ensure global interconnectedness. The transnational comparative analysis, based on linear regression between node and edge cardinality, places Italy above the average European trend, characterizing it as a case of a hyper-collaborative, non-scalable ecosystem. The results confirm the utility of network metrics for dynamically modeling the production structure of a cultural sector. Qualitative analysis of the quantitative findings of the study.

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