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?

2. How does the Italian scene position itself relative to major European and non-European countries?

3. Who are the most central and influential artists, and what structural role do they occupy?

4. How are connections distributed within the network?

5. What are the prevalent collaboration patterns? Do artists tend to collaborate with peers similar in terms of connectivity, popularity, or musical genre?
6. Do well-defined communities exist in the network? How do they relate to musical genres?
7. What collaborative strategies do emerging artists adopt compared to established artists?

To address these questions, the study applies a wide range of network metrics and analysis techniques to a dataset derived from Spotify, appropriately enriched with information on artist nationality and genre.

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 directed 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 (according to kworb.net)
- **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 the artists in the dataset, a procedure was designed with the objective of maximizing the number of artists assigned a musical genre, while simultaneously maintaining the consistency and semantic plausibility of the assignments. The process unfolds across multiple successive stages, each intervening only on artists still lacking a genre from the previous step.

- **Direct mapping of Spotify genres.** In a first stage, the specific genres provided by Spotify (e.g., *italian pop*, *alternative rock*, *deep house*) were normalized and mapped to a limited set of musical *macro-categories* (such as *Pop*, *Rock*, *Hip Hop / Rap*, *Electronic / Dance*, etc.). This mapping was implemented through lexical and keyword-based rules, enabling the reduction of the high fragmentation of the original genres and yielding a more compact and comparable representation.

- **AI-assisted completion.** Genres that could not be automatically mapped in the previous stage (collected in the *Others / Specific* category) were extracted and provided as input to an artificial intelligence-assisted classification process. The AI mapped each remaining genre to one of the previously defined musical macro-categories, based on semantic similarities and general musical knowledge. The associations obtained were subsequently reintegrated into the dataset, allowing for a significant reduction in the number of unclassified genres and an improvement in the overall mapping coverage, while maintaining consistency with the adopted categorization scheme.

- **Inference via artistic collaborations.** For artists still lacking a genre after direct mapping, the collaboration graph was leveraged. Specifically, the most frequent genres among their direct collaborators were assigned; if this was insufficient, the inference was extended through a Breadth-First Search (BFS) up to three levels of distance in the network, selecting the most recurrent genres among the visited nodes.

- **Inference based on popularity metrics.** Artists still unclassified were analyzed based on quantitative indicators such as follower count and popularity. Through simple heuristics derived from patterns observed in the dataset (e.g., high popularity associated with mainstream genres), the most probable genres were inferred.

- **Global fallback assignment.** Finally, for the rare cases still without a genre, a fallback mechanism based on the globally most common genres in the dataset was applied, ensuring that every artist was associated with at least one musical macro-category.

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 initial dataset is derived from the Spotify API and pre-processed to explicitly model collaboration relationships among artists, providing a reasonable approximation of actual musical interactions on the platform. However, the dataset reflects Spotify's ecosystem and temporal snapshot, and therefore may not capture collaborations occurring outside the platform or informal artistic relationships.
Additional artist-level attributes, such as nationality and dominant musical genre, were introduced to support higher-level social and cultural analyses. Nationality was inferred using a hybrid approach combining genre-based cues and external data from MusicBrainz, applied selectively to reduce ambiguity due to artist name homonyms. While this procedure increases coverage and interpretive power, it introduces a degree of uncertainty, particularly for artists whose identity or geographic origin is weakly signaled by available metadata.
Similarly, musical genres were consolidated into a limited set of macro-categories through a multi-step process involving rule-based mapping, AI-assisted classification, and network-based inference. This approach improves comparability and completeness, but necessarily abstracts away finer-grained genre distinctions and may propagate local biases through collaboration-based inference.
Regarding reliability, all data processing and enrichment steps follow deterministic rules or documented heuristics, ensuring that the analysis is reproducible given the same inputs and parameters. External data sources and AI-assisted mappings represent potential sources of variability; however, their use was constrained to well-defined stages and applied consistently across the dataset. Overall, the adopted methodology yields a dataset that is both valid and reliable for social network analysis, as it is grounded in authoritative data sources, enriched through controlled and well-documented procedures, and constructed to balance descriptive accuracy with reproducibility, making it a solid foundation for the subsequent network and community analyses.

4 Measures and Results

This section concisely summarizes the main measures used, the technologies employed, and their connection to the study's objectives.

Network Representation

- Undirected graph $G = (V, E)$: nodes = artists (`spotify_id`), edges = collaborations between artists present in tracks.

- Python implementation using `pandas` for node/edge CSV files and `NetworkX` for graph construction and measure calculation.

Centrality Measures

- **Degree centrality**: normalizes the number of collaborations for each artist, identifies the most connected hubs, and is used to select the top artists in the analysis subgraph.
- **Betweenness centrality**: measures how often an artist lies on the shortest paths between pairs of nodes, identifying structural "brokers" between different communities and genres.
- **Closeness centrality**: inverse of the average distance from an artist to all others, quantifies how quickly an artist can reach the rest of the network.
- **Eigenvector centrality**: assigns higher scores to artists connected to other central artists, capturing membership in the scene's "core".

Community Detection and Bridges

- **Louvain**: identifies communities by maximizing modularity, allowing the association of structural clusters with macro-genres, national scenes, or label groups.
- **Edge betweenness** and **Burt's constraint**: respectively identify bridge collaborations between communities and artists with access to *structural holes*, fundamental for the diffusion of styles and content between different worlds.

Genres, Nationality, and Success

- Genres and nationalities are managed as node attributes (`genre`, `nationality`); intra/inter-genre and intra/inter-nationality collaborations are counted to assess assortativity and transnational openness.
- For artists without a genre, the genre is inferred from the most frequent genre in their network neighborhood, with a minimum collaboration threshold to ensure robustness.
- Structural measures are correlated with external indicators (Spotify `popularity`, number of collaborations, foreign collaborations, chart presence) to study the link between network position, popularity, and international expansion.

Emerging Artists

- A `DataFrame` is constructed with `popularity` and number of collaborations per artist; thresholds on both indicators define three classes: *emerging*, *intermediate*, *established*.
- The collaboration matrix between classes (emerging–emerging, emerging–established, etc.) shows networking strategies (horizontal among peers vs. linking to established artists) and how these are reflected in growth in centrality and popularity.

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 niches or artistic communities with limited contact with the rest of the national ecosystem. Each artist is connected, on average, 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 realized) 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.

4.2 Comparative Analysis with Major European Countries

To situate the Italian results in a broader continental context, the analysis was extended to major European countries. This approach enables a comparative evaluation of collaborative dynamics.
Table 2 shows the top five European countries by number of artists (Italy, France, Germany, United Kingdom, and the Netherlands) and, subsequently, other analyzed countries, providing a comprehensive picture of the main European collaborative networks.

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 between countries, highlighting distinct models of musical collaboration:

1. **Netherlands**: They have the highest **average degree** in Europe (7.24) despite ranking only fifth in 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.

2. **Poland**: With an **average degree** of 6.20 and an **average clustering coefficient** of 0.167, they represent a model of a highly cohesive and interconnected network. Only 8 **connected components** indicate excellent structural integration.

3. **Greece**: Has the highest **average clustering coefficient** in Europe (0.183) and a very good **average degree** (5.47). With only 4 **connected components**, it is one of the best-integrated networks.

4. **United Kingdom**: Despite its larger size (3,290 artists), it 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). This difference reflects two distinct structural configurations: on one hand, a British network that is extensive but characterized by high dispersion of collaborative relationships; on the other, an Italian ecosystem of more contained dimensions but relatively denser and more cohesive.

4.2.1 Multidimensional Analysis and Interpretation of European Patterns

Figure 1 presents a comparative multidimensional analysis of musical collaboration networks at the European level. The analysis synthesizes 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 center, the total number of collaborations. In the latter, we observe:

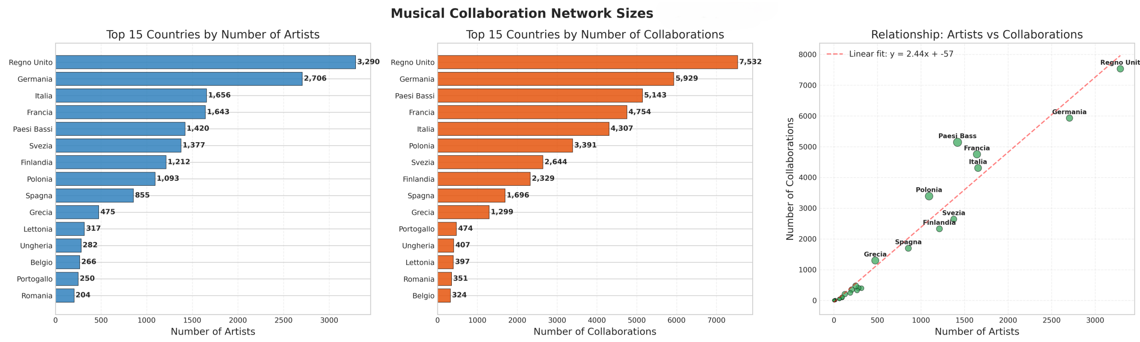


Figure 1: Comparative multidimensional analysis of musical collaboration networks at the European level.

- 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 for 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).

4.2.2 Structural Relationship Between Artists and Collaborations

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 analyzed using linear regression estimated by the least squares method, considering all European countries included in the study. The resulting line, $y = 2.44x - 76$, describes the average trend of European musical networks.

- **Slope coefficient (2.44)**: indicates that, on average, for each additional artist, there are approximately 2.44 more collaborations. This value expresses the average collaborative intensity of the analyzed musical networks.
- **Intercept (-76)**: the negative value suggests that very small networks tend to have a number of collaborations lower than that predicted by the general trend. In particular, below approximately 30 artists, the network structure appears less active.

The regression line can therefore be used as a reference: countries positioned above it show a level of collaboration higher than expected given their size, while those below exhibit more contained collaborative activity.

- Countries positioned **above the regression line** (Netherlands, Poland, Greece) represent hyper-collaborative ecosystems, where the volume of interactions systematically exceeds expectations given their dimensions. These systems are typically characterized by high values of density and clustering coefficient.
- Countries **close to the line** (Italy, France, Germany, United Kingdom) follow an approximately linear relationship between collaborative activity and size.

- The **dispersion** of the data confirms the absence of a single European model, highlighting instead a plurality of structural configurations.

4.3 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 analysis reveals a marked dimensional asymmetry between the analyzed musical systems:

- The **United States** present exceptional dimensional characteristics, with a network of 6,217 artists and 14,860 collaborations, over 3.7 times larger than the Italian network. This considerable size is accompanied by a low density (0.000769) that reflects the large scale and diversification of the US music market, with 111 distinct groups operating in a relatively isolated manner. The average degree (4.78) is slightly lower than the Italian figure (5.20), indicating that, despite the larger size, the US network shows lower intensity of connections per artist.

- **Brazil** (1,859 artists) and **India** (999 artists) represent the major emerging musical systems, both characterized by high-density networks (0.003212 and 0.005685 respectively) and high structural cohesion. These values indicate highly integrated musical ecosystems, with only 3 and 2 distinct groups respectively, reflecting a strong cultural and geographical unity.

- Italy ranks sixth among non-European countries globally for size, surpassing countries such as South Korea (708) and Australia (653). Compared to Japan (890), Italy shows a much more cohesive structure: Japan, despite having more artists, presents a more fragmented network (38 distinct groups) and a lower average degree (2.48), while Italy, with 1,656 artists, maintains strong integration (16 components) and high collaborative activity (average degree 5.20).

The analysis identifies several models of musical ecosystems characterized by intense collaborative activity:

- **Puerto Rico** represents the most pronounced case of hyper-collaboration, with a particularly high average degree (10.66) and a high clustering coefficient (0.314).

- **Brazil** and **India** show similar structures, with high average degrees (5.97 and 5.67 respectively) and significant clustering (0.173 and 0.181). These values suggest vibrant ecosystems with a strong tendency towards the formation of cohesive communities, supported by low fragmentation (3 and 2 distinct groups), which indicates unified and well-integrated music scenes.

- **South Korea** presents an interesting case of a moderately large network (708 artists) but with high density (0.006185) and significant average degree (4.37), reflecting a structured and interconnected music scene.
- Italy, with an average degree of 5.20, ranks among the countries with the highest collaborative activity globally, surpassing countries such as the United States (4.78), Germany (4.38), the United Kingdom (4.58), and South Korea (4.37). This positioning indicates a musical culture particularly oriented towards collaboration, which compensates for the more contained network size with a greater intensity of interactions.

4.3.1 Density and Fragmentation

- The analyzed **African** countries (Ghana, Egypt, Nigeria, Algeria) represent the model of small, ultra-dense networks, with densities ranging from 0.047 to 0.143. These networks, despite having a limited number of artists (14-66 nodes), show extremely high internal interconnection. In particular, Ghana (0.138) and Algeria (0.143) present densities over 40 times higher than the US figure (0.000769), indicating extremely compact music scenes where almost all artists collaborate directly with each other.
- **Latin American** networks show intermediate densities but significantly higher than European and North American networks: Venezuela (0.077), Panama (0.065), Colombia (0.010), Argentina (0.010). These values, ranging from 3 to 10 times the Italian density, reflect regionally cohesive music scenes that are sufficiently broad to support a certain internal diversification. The low fragmentation (2-5 connected components) confirms the high degree of integration of these networks.
- Italy (density 0.00314) occupies an intermediate position in the global panorama. When compared to the top 5 non-European countries by size, it presents a density:
  - 4 times higher than that of the United States (0.000769)
  - similar to that of Brazil (0.003212)
  - about half that of India (0.005685) and Mexico (0.005580)
  - higher than that of Japan (0.002788)

  This positioning reflects a network that effectively balances size (1,656 nodes) and cohesion.
- Among countries with networks of similar size to Italy (1,000-2,000 nodes), India (999 nodes, density 0.005685) and South Korea (708 nodes, density 0.006185) show densities almost double that of Italy, reflecting more intense collaboration models. However, Italy compensates with an average degree (5.20) higher than that of many of these countries, indicating that despite a moderate overall density, Italian artists tend to collaborate with a greater number of colleagues.

4.3.2 Clustering and Local Cohesion

- The highest values of the **clustering coefficient** are observed in **Puerto Rico** (0.314), **Ghana** (0.287), and the **Dominican Republic** (0.275). These exceptionally high values, ranging from 8 to 11 times those of countries like Canada or Australia, indicate a strong tendency towards the formation of "closed triads" in which an artist's collaborators tend to frequently collaborate with each other as well. This model suggests extremely cohesive musical communities.

- **India** (0.181) and **Brazil** (0.173) show high clustering values which, combined with the significant sizes of their networks (999 and 1,859 nodes respectively), reflect music scenes that combine breadth with strong internal cohesion. These values, approximately 1.5-2 times higher than the Italian one, indicate musical ecosystems in which the formation of closely interconnected artistic communities is noted.
- Italy (0.119) presents a moderate but significant clustering value in the global context. The comparison with major non-European countries reveals that:
  - Italian clustering is higher than that of the United States (0.082) and Japan (0.042)
  - It is lower than that of India (0.181), Brazil (0.173), and Mexico (0.156)
  - It occupies an intermediate position among countries with networks of similar dimensions

  This positioning indicates a network that balances local cohesion and openness: sufficiently cohesive to foster the formation of stable artistic circles, but also open enough to allow new connections and external exchanges.
- The low clustering values in countries such as **Canada** (0.037), **Australia** (0.035), and **China** (0.027) suggest network structures less inclined towards triadic closure. These values are approximately 3-4 times lower than the Italian one.

4.4 Nodes

4.4.1 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.

4.4.1.1 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:
Guè emerges as the most connected node in the network with 114 distinct collaborations, establishing himself as a central hub of the Italian music scene. The presence of Andrea Bocelli in second place with 103 collaborations is particularly significant: despite working in a substantially different musical genre (classical/pop crossover), he has developed an extensive network

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
4	Gemitaiz	0.0489
5	Night Skinny	0.0483
6	Don Joe	0.0477
7	Inoki	0.0441
8	Fabri Fibra	0.0435
9	Emis Killa	0.0429
10	Ennio Morricone	0.0411

of collaborations spanning multiple genres. The predominance of hip hop artists in the top positions (Clementino, Gemitaiz, Night Skinny, Don Joe, Inoki, Fabri Fibra, Emis Killa) confirms that this genre has a high propensity for collaboration and constitutes a central element in the network's structure. The presence of Ennio Morricone in tenth place introduces an element of generational and stylistic diversity into the ranking.

4.4.1.2 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 an even more pronounced concentration compared to degree centrality: the average value is 0.0089 while the maximum reaches 0.2573, highlighting that a small number of artists form a highly cohesive central core.
Table 5 presents the ten artists with the highest eigenvector centrality:

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
4	Night Skinny	0.1903
5	Fabri Fibra	0.1717
6	Marracash	0.1677
7	Jake La Furia	0.1664
8	Don Joe	0.1650
9	MadMan	0.1533
10	Lazza	0.1425

Guè maintains the dominant position with a value considerably higher than the other artists (approximately 25% more than the second-ranked artist), indicating that his collaborations predominantly involve the most central artists in the scene. The exclusive presence of hip hop

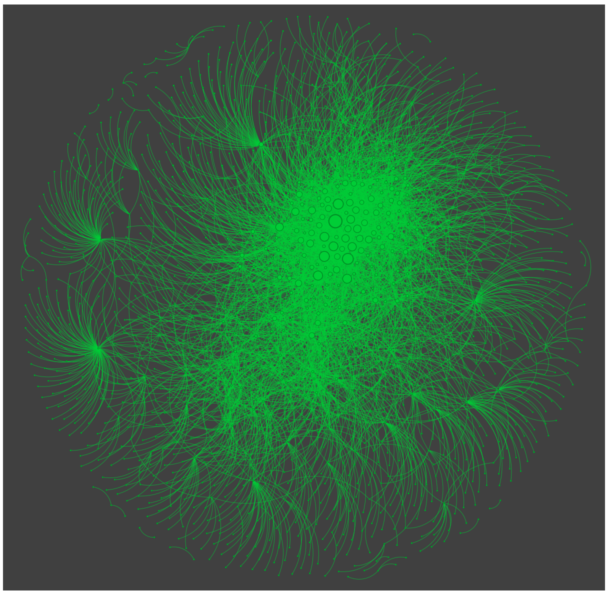


Figure 2: Caption

artists in this ranking reveals the existence of a central core dominated by this genre. Artists such as Marracash, Jake La Furia, Don Joe, MadMan, and Lazza form a highly interconnected core that defines the center of the Italian hip hop network. It is noteworthy that Andrea Bocelli, despite having a high number of collaborations (second in degree centrality), is absent from this ranking, suggesting that his collaborations predominantly involve less central or more peripheral artists compared to the network's main core.

4.4.1.3 Closeness Centrality

Closeness centrality measures the closeness of a node to all other nodes in the network, calculating the inverse of the average distance based on shortest paths. An artist with high closeness centrality can rapidly reach any other artist in the network through a limited number of intermediaries.
The results show an average value of 0.2336 with a maximum of 0.3677, indicating that even the most central artists require an average of about three steps to reach any other node in the network.
In the practical context of the music scene, high closeness centrality confers concrete strategic advantages: optimally positioned artists can more easily access information and emerging trends in different parts of the scene, initiate collaborations with distant artists in the network through few intermediaries, and rapidly disseminate their own style or musical innovations throughout the entire ecosystem. This metric therefore identifies artists who, regardless of their belonging to the central core, occupy positions that facilitate communication and the circulation of opportunities throughout the entire network.
Table 6 presents the ten artists with the highest closeness centrality:
Guè confirms his dominant position, being optimally positioned to access the entire network with the minimum number of intermediaries. Clementino and J-AX emerge particularly in this metric compared to eigenvector centrality, suggesting they occupy strategic bridging positions: although not in the densest core, their collaborations traverse different network clusters, allowing them to act as efficient connectors between different subcommunities. The presence of Elisa (ninth position) and Rocco Hunt (tenth position) introduces significant genre diversifica-

Table 6: Top 10 artists by Closeness Centrality

Rank	Artist	Closeness Centrality
1	Guè	0.3677
2	Clementino	0.3576
3	Gemitaiz	0.3537
4	Fabri Fibra	0.3500
5	J-AX	0.3487
6	Night Skinny	0.3480
7	Marracash	0.3466
8	Emis Killa	0.3464
9	Elisa	0.3447
10	Rocco Hunt	0.3420

tion. These artists, although not belonging to the dominant hip hop core, maintain positions of global closeness that facilitate the transmission of cross-genre influences. The network, although dominated by hip hop in its central core, therefore maintains a structure that allows efficient communication between different genres, suggesting a musical ecosystem where barriers between genre communities do not hinder the rapid circulation of ideas and collaborative opportunities.

4.4.1.4 Betweenness Centrality

Betweenness centrality identifies nodes that are frequently found on the shortest paths between other pairs of nodes. Artists with high betweenness act as intermediaries between different components of the network, even without necessarily having a high number of direct connections. The calculation was approximated by sampling 1000 random nodes for computational efficiency reasons.

The results show a strongly asymmetric distribution: the average value is 0.0018 while the maximum reaches 0.1250, indicating that a very limited number of nodes controls the flow of connections between different clusters in the network.

Andrea Bocelli dominates this metric with a value significantly higher than the other artists (almost 40% higher than the second-ranked artist), revealing his structural role as a bridge between different musical contexts. Despite not belonging to the hip hop core identified by eigenvector centrality, Bocelli connects classical and pop music with other genres, acting as an essential intermediary in the overall network structure. The presence of Ennio Morricone (fifth position), DJ Matrix (sixth position), Jovanotti (ninth position) and Cristina D'Avena (tenth position) - artists working in genres other than hip hop - confirms that betweenness centrality captures a structural role distinct from local centrality. These artists do not belong to the central core but occupy gatekeeping positions between different genre communities. Clementino maintains a high position in all analyzed metrics (third in degree, second in closeness and betweenness), confirming himself as one of the structurally most relevant artists in the network: an integral part of the hip hop core but with strong bridging capabilities towards other music scenes.

Table 7 presents 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	Elisa	0.0611
5	Ennio Morricone	0.0575
6	DJ Matrix	0.0558
7	J-AX	0.0538
8	Inoki	0.0509
9	Jovanotti	0.0452
10	Cristina D'Avena	0.0422

4.4.2 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.

4.4.2.1 Degree Assortativity

The obtained degree assortativity coefficient is **-0.1052**, indicating a slightly **disassortative** network. This negative value means that artists with numerous collaborations tend to connect with artists who have fewer, rather than collaborating predominantly with each other.
In the context of the analyzed musical network, this pattern suggests that hubs (artists with high degree) do not form an isolated group but include less prolific artists in their collaborations. This may reflect various dynamics: established artists providing visibility to emerging artists, producers and featuring artists collaborating with a broad spectrum of artists at different activity levels, or more generally a relatively open music scene where the number of past collaborations does not constitute a significant barrier for future opportunities. The negative value, albeit moderate, is consistent with a structure that facilitates mobility and access even for less central artists, in contrast with rigidly hierarchical models where exclusively the most successful figures collaborate with each other.

4.4.2.2 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 result suggests that popularity, measured through the number of Spotify followers, exerts a limited influence on collaboration patterns. Artists with a high number of followers do not show a marked preference for exclusive collaborations with other artists of similar popularity, nor does a clear segregation emerge between different popularity strata. Detailed analysis categorizes edges into High-High, Low-Low and High-Low using the median as a threshold, revealing that there are both collaborations between artists of similar popularity and collaborations with artists of a different popularity level from their own, in relatively balanced proportions.
This pattern indicates a relatively fluid Italian musical ecosystem regarding fame: collaborations are not strongly constrained by the number of followers, and established artists show

willingness to collaborate with less-known colleagues. However, the slightly positive value suggests the existence of a weak homophilic preference: very popular artists collaborate more frequently with other popular artists, but this tendency does not create significant structural barriers.

4.4.2.3 Genre Assortativity and Modularity

The results regarding musical genre show markedly different patterns compared to the other forms of assortativity. The genre assortativity coefficient is **0.4778**, a decidedly positive value indicating a strong assortative tendency. The modularity is **Q = 0.2987**, classifiable as moderate according to the implemented interpretative scale (between 0.1 and 0.3).
These values reveal that musical genre exerts a significant influence on collaboration patterns: artists tend to collaborate predominantly with other artists of the same genre or related genres. However, the moderate (not high) modularity indicates that this tendency does not produce impermeable compartments: there are numerous cross-genre collaborations that keep the network globally interconnected.
Table 8 presents the ten most frequent genre pairs in collaborations:

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

Hip Hop / Rap dominates with 1829 intra-genre collaborations, confirming the high density and cohesion of this musical community. Pop presents 819 internal collaborations, also demonstrating significant internal cohesion. The Hip Hop / Rap + Pop combination accounts for 596 edges, constituting by far the most frequent cross-genre contamination. This reveals a fundamental structural bridge between the two dominant genres of the Italian scene. The Hip Hop / Rap + Indie (104 edges) and Elettronica / Dance + Pop (62 edges) combinations represent other significant contaminations, although with considerably lower frequency. Genres such as Classica/Orchestrale (72 internal edges), Elettronica/Dance (85 edges), and Soundtrack (37 edges) highlight communities of more limited size but still internally cohesive.
The ratio between intra-genre and inter-genre edges, calculated by the function through the analysis of overlaps between genre lists, shows that collaborations remain predominantly within genre boundaries, with relevant exceptions.
In the context of the Italian music scene, these results describe a picture of specialization with selective contamination: genres maintain distinct identities and artists collaborate predominantly within their stylistic boundaries, but there are consolidated bridges between complementary genres, particularly between hip hop and pop. The high density of hip hop (1829 internal

collaborations) confirms what emerged from the centrality analysis: this genre not only dominates the central core of the network but has developed an extremely rich and interconnected internal ecosystem.

4.5 Network

4.5.1 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.

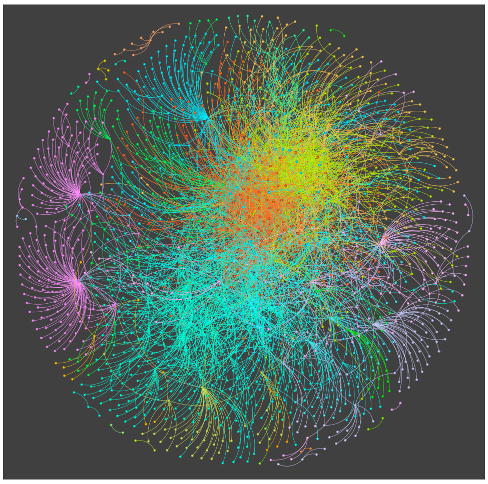


Figure 3: Community structure obtained by applying the Louvain algorithm in Gephi.

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.

4.5.2 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.

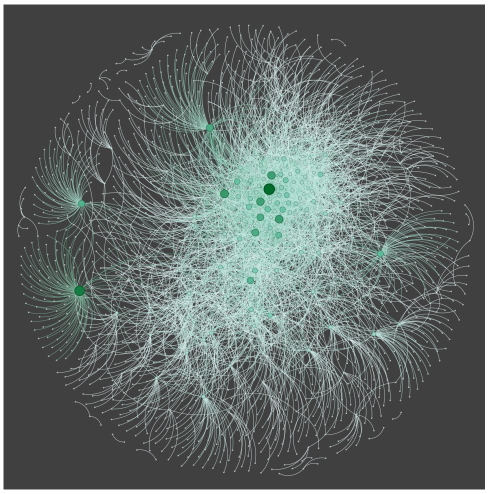
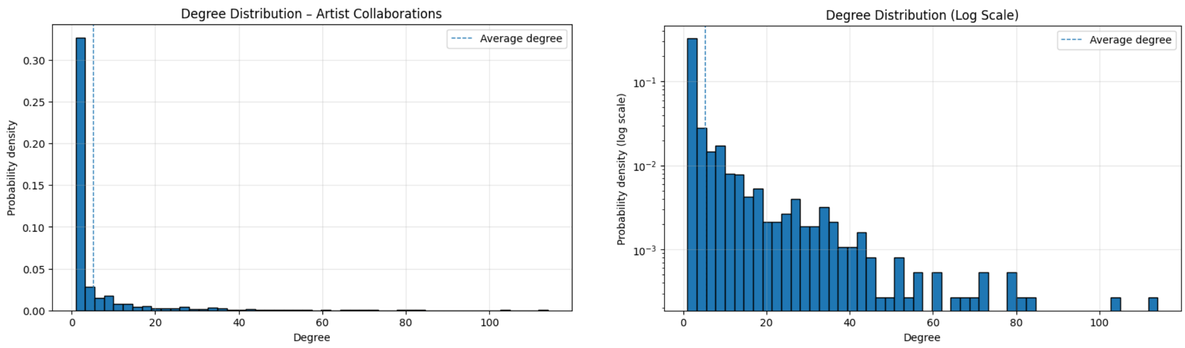


Figure 4: Artist collaboration network visualized in Gephi with node size proportional to degree.



(a) Degree distribution in linear scale.

(b) Degree distribution in logarithmic scale.

Figure 5: Degree distribution of the artist collaboration network. The linear-scale histogram highlights the high concentration of low-degree nodes, while the logarithmic-scale representation emphasizes the long-tailed behavior induced by a small number of highly connected artists.

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