

Analisi delle Collaborazioni Musicali su Spotify: Una Prospettiva di Social Network Analysis

Federico , matricola
Federica Santisi, matricola
Giorgia Pirelli, matricola

Dicembre 2025

1 Introduction

L'industria musicale ha subito trasformazioni radicali negli ultimi due decenni, passando da un modello tradizionale basato sulla vendita fisica di dischi a un ecosistema digitale dominato dalle piattaforme di streaming. Spotify, lanciato nel 2008, è diventato il principale servizio di streaming musicale a livello globale, con oltre 500 milioni di utenti attivi e un catalogo che supera i 100 milioni di brani.

In questo nuovo panorama, le collaborazioni musicali, comunemente note come "featuring", hanno assunto un ruolo centrale. Se in passato le collaborazioni erano eventi relativamente rari e spesso limitati a progetti speciali, oggi rappresentano una strategia fondamentale per artisti di ogni livello di popolarità. Le collaborazioni permettono agli artisti di raggiungere nuove audience, sperimentare con generi musicali diversi, aumentare la propria visibilità sulle piattaforme di streaming e creare connessioni strategiche all'interno dell'industria musicale.

La Social Network Analysis (SNA) offre strumenti metodologici potenti per studiare queste dinamiche relazionali. Rappresentando gli artisti come nodi e le loro collaborazioni come archi, è possibile costruire una rete che cattura la complessità delle interazioni nel panorama musicale contemporaneo.

Questo studio si concentra sull'analisi delle collaborazioni tra artisti su Spotify, con un focus particolare sulla scena musicale italiana come caso di studio principale.

Domande di Ricerca

Attraverso una prospettiva di rete, questo lavoro mira a rispondere a domande fondamentali sul funzionamento dell'ecosistema musicale italiano e il suo posizionamento nel contesto internazionale:

1. Quali sono le caratteristiche strutturali fondamentali della rete di collaborazioni musicali italiane su Spotify?
2. Come si posiziona la scena italiana rispetto ai principali paesi europei ed non europei?
3. Chi sono gli artisti più centrali e influenti, e che tipo di ruolo strutturale ricoprono?
4. Come sono distribuiti i collegamenti nella rete?

5. Quali sono i pattern di collaborazione prevalenti? Gli artisti tendono a collaborare con colleghi simili per grado di connessione, popolarità o genere musicale?
6. Esistono comunità ben definite nella rete? Come si relazionano con i generi musicali?
7. Quali strategie collaborative adottano gli artisti emergenti rispetto agli artisti affermati?

Per rispondere a questi interrogativi, lo studio applica un'ampia gamma di metriche e tecniche di analisi di rete a un dataset ricavato da Spotify, opportunamente arricchito con informazioni su nazionalità e genere degli artisti.

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 Aggiunta del genere artistico

Per associare uno o più generi musicali agli artisti del dataset, è stata progettata una procedura con l'obiettivo di massimizzare il numero di artisti con genere musicale assegnato, mantenendo al contempo coerenza e plausibilità semantica delle assegnazioni. Il processo si articola in più fasi successive, ciascuna delle quali interviene solo sugli artisti rimasti privi di genere nel passaggio precedente.

- **Mappatura diretta dei generi Spotify.** In una prima fase, i generi specifici forniti da Spotify (ad esempio *italian pop*, *alternative rock*, *deep house*) sono stati normalizzati e ricondotti a un insieme limitato di *macro-categorie* musicali (come *Pop*, *Rock*, *Hip Hop / Rap*, *Elettronica / Dance*, ecc.). Questa mappatura è stata realizzata tramite regole lessicali e keyword-based, consentendo di ridurre l'elevata frammentazione dei generi originali e di ottenere una rappresentazione più compatta e comparabile.
- **Completamento assistito tramite AI.** I generi che non risultavano mappabili automaticamente nella fase precedente (raccolti nella categoria *Altri / Specifici*) sono stati estratti e forniti come input a un processo di classificazione assistito da intelligenza artificiale. L'AI ha ricondotto ciascun genere residuo a una delle macro-categorie musicali precedentemente definite, sulla base di similarità semantiche e conoscenza musicale generale. Le associazioni così ottenute sono state successivamente reintegrate nel dataset, consentendo di ridurre in modo significativo il numero di generi non classificati e di migliorare la copertura complessiva della mappatura, mantenendo coerenza con lo schema di categorizzazione adottato.
- **Inferenza tramite collaborazioni artistiche.** Per gli artisti privi di genere dopo la mappatura diretta, è stato sfruttato il grafo delle collaborazioni. In particolare, sono stati assegnati i generi più frequenti tra i collaboratori diretti; qualora ciò non fosse sufficiente, l'inferenza è stata estesa tramite una ricerca BFS (Breadth-First Search) fino a tre livelli di distanza nella rete, selezionando i generi più ricorrenti nei nodi visitati.
- **Inferenza basata su metriche di popolarità.** Gli artisti ancora non classificati sono stati analizzati in base a indicatori quantitativi come numero di follower e popolarità. Attraverso semplici euristiche derivate da pattern osservati nel dataset (ad esempio alta popolarità associata a generi mainstream), sono stati inferiti i generi più probabili.
- **Assegnazione di fallback globale.** Infine, per i rari casi rimasti senza genere, è stato applicato un meccanismo di fallback basato sui generi globalmente più comuni nel dataset, garantendo che ogni artista fosse associato ad almeno una macro-categoria musicale.

Questa strategia multilivello ha permesso di ottenere un dataset completo e consistente dal punto di vista dei generi musicali, riducendo al minimo le assegnazioni arbitrarie e sfruttando in modo integrato informazioni semantiche, strutturali e quantitative. Il risultato finale costituisce una base solida per le successive analisi di rete e di comunità artistiche.

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 Misure e Risultati

In questa sezione si riassumono in modo sintetico le principali misure utilizzate, le tecnologie impiegate e il loro legame con gli obiettivi dello studio.

Rappresentazione della rete

- Grafo non orientato $G = (V, E)$: nodi = artisti (spotify_id), archi = collaborazioni tra artisti presenti nelle tracce.
- Implementazione in Python con pandas per i CSV dei nodi/archi e NetworkX per la costruzione del grafo e il calcolo delle misure.

Misure di centralità

- **Degree centrality:** normalizza il numero di collaborazioni di ciascun artista, identifica gli hub più connessi e viene usata per selezionare i top artisti nel sottografo di analisi.
- **Betweenness centrality:** misura quante volte un artista cade sui cammini minimi tra coppie di nodi, individuando i “broker” strutturali tra comunità e generi diversi.
- **Closeness centrality:** inverso della distanza media da un artista a tutti gli altri, quantifica quanto rapidamente un artista può raggiungere il resto della rete.
- **Eigenvector centrality:** assegna punteggi più alti agli artisti collegati ad altri artisti centrali, catturando l'appartenenza al “core” della scena.

Community detection e bridge

- **Louvain:** individua comunità massimizzando la modularità, permettendo di associare cluster strutturali a macro-generi, scene nazionali o gruppi di etichetta.
- **Edge betweenness e constraint** di Burt: identificano rispettivamente collaborazioni-ponte tra comunità e artisti con accesso a *structural holes*, fondamentali per la diffusione di stili e contenuti tra mondi diversi.

Generi, nazionalità e successo

- Generi e nazionalità sono gestiti come attributi dei nodi (genre, nationality); si contano collaborazioni intra/inter-genere e intra/inter-nazionali per valutare assortatività e aperture transnazionali.
- Per gli artisti senza genere, il genere viene inferito dal genere più frequente nel vicinato di rete, con soglia minima di collaborazioni per garantire robustezza.
- Le misure strutturali sono correlate con indicatori esterni (popularity Spotify, numero collaborazioni, collaborazioni estere, presenza in chart) per studiare il legame tra posizione nella rete, popolarità ed espansione internazionale.

Artisti emergenti

- Si costruisce un DataFrame con popularity e numero di collaborazioni per artista; soglie su entrambi gli indicatori definiscono tre classi: *emergente*, *intermedio*, *affermato*.
- La matrice delle collaborazioni tra classi (emergente-emergente, emergente-affermato, ecc.) mostra le strategie di networking (orizzontale tra pari vs collegamento verso artisti affermati) e come queste si riflettano nella crescita di centralità e popolarità.

4.1 Analisi generale della rete di collaborazioni musicali italiane

Al fine di delineare le differenze strutturali chiave e ottenere una panoramica della topologia e delle dinamiche interne della rete di collaborazioni musicali italiane, è stata calcolata una serie di misure generali, come riportato nella Tabella 1.

Il **numero di componenti connesse** è stato misurato per identificare la presenza di sottogruppi isolati all'interno dell'ecosistema musicale nazionale. Sia il **grado massimo** che il **grado medio** dei nodi sono stati calcolati per valutare la connettività globale della rete e per quantificare l'intensità dell'attività collaborativa degli artisti. La **densità** della rete è stata calcolata per misurare la sua coesione complessiva. Il **diametro** e la **lunghezza media del percorso più breve** forniscono una valutazione dell'efficienza del flusso di informazioni e indicano la facilità con cui gli artisti possono essere raggiunti gli uni attraverso gli altri. Infine, il **coefficiente di clustering medio** e la **transitività** sono stati calcolati per misurare la coesione locale della rete e la tendenza alla formazione di gruppi strettamente connessi.

Table 1: Caratteristiche strutturali della rete di collaborazioni musicali italiane

Parametro	Valore
Nodi totali (artisti)	1.656
Archi totali (collaborazioni)	4.307
Componenti connesse	16
Grado massimo del nodo	114
Grado medio del nodo	5.20
Densità	0.00314
Diametro	10
Lunghezza media del percorso più breve	4.14
Coefficiente di clustering medio	0.119
Transitività	0.128

I sedici componenti connessi suggeriscono l'esistenza di gruppi isolati, corrispondenti a nicchie o comunità artistiche con contatti limitati con il resto dell'ecosistema nazionale. Ciascun artista

è connesso, in media, a circa cinque colleghi nella rete. Tuttavia, la distribuzione del grado è fortemente eterogenea: la presenza di un nodo con grado 114 rivela un *hub* di importanza centrale, mentre il 54.6% degli artisti presenta un grado pari a 1, collaborando dunque con un solo altro artista. La densità estremamente bassa (circa lo 0.31% delle possibili connessioni è realizzata) conferma la natura **sparsa** della rete. Il diametro della rete è pari a 10, indicando che due artisti qualsiasi possono essere collegati attraverso al massimo 10 collaborazioni intermedie. Nonostante la bassa densità, la rete presenta una lunghezza media del percorso contenuta (4.14), indicando che gli artisti sono collegati attraverso poche collaborazioni intermedie.

I valori del coefficiente di clustering medio (0.119) e della transitività (0.128) sono moderati e tra loro vicini. Ciò indica una tendenza misurabile, seppur non dominante, alla **chiusura triadica**: due collaboratori di uno stesso artista hanno una probabilità di circa il 12% di aver collaborato a loro volta. Questa coesione locale favorisce la formazione di circoli artistici coesi e comunità parzialmente clusterizzate, contribuendo alla stabilità delle relazioni collaborative e alla condivisione di pratiche artistiche all'interno di sottogruppi, pur mantenendo sufficiente apertura per permettere connessioni tra comunità diverse.

4.2 Analisi comparativa con i principali paesi europei

Per collocare i risultati italiani in un contesto continentale più ampio, l'analisi è stata estesa ai principali paesi europei. Questo approccio consente una valutazione comparativa delle dinamiche collaborative.

La Tabella 2 mostra i primi cinque paesi europei per numero di artisti (Italia, Francia, Germania, Regno Unito e Paesi Bassi) e, successivamente, altri paesi analizzati, fornendo un quadro

completo delle principali reti collaborative europee.

Table 2: Confronto con i principali paesi europei

Metrica	Italia	Francia	Germania	Regno Unito	Paesi Bassi
Total nodes	1,656 (3°)	1,643 (4°)	2,706 (2°)	3,290 (1°)	1,420 (5°)
Total edges	4,307 (5°)	4,754 (4°)	5,929 (2°)	7,532 (1°)	5,143 (3°)
Average degree	5.20 (5°)	5.79 (2°)	4.38 (7°)	4.58 (6°)	7.24 (1°)
Connected components	16 (4°)	27 (6°)	32 (7°)	70 (9°)	12 (3°)
Average clustering	0.119 (4°)	0.113 (6°)	0.120 (3°)	0.062 (9°)	0.151 (2°)

Dall’analisi emergono differenze strutturali significative tra i paesi, evidenziando modelli distinti di collaborazione musicale:

- Paesi Bassi:** Presentano l’**average degree** più alto d’Europa (7.24) nonostante siano solo quinti per numero di artisti. Il loro **average clustering coefficient** (0.151) è il secondo più alto, e la percentuale di hub artists (17.8%) è la massima in Europa.
- Polonia:** Con un **average degree** di 6.20 e **average clustering coefficient** di 0.167, rappresentano un modello di rete altamente coesa e interconnessa. Solo 8 **connected components** indicano un’eccellente integrazione strutturale.
- Grecia:** Ha il **average clustering coefficient** più alto d’Europa (0.183) e un ottimo **average degree** (5.47). Con soli 4 **connected components**, è una delle reti meglio integrate.
- Regno Unito:** Nonostante le dimensioni maggiori (3.290 artisti), presenta il **average clustering coefficient** più basso (0.062) e la frammentazione più alta (70 **connected components**), riflettendo un mercato vasto ma segmentato. Presenta anche una densità di rete significativamente più bassa (0.001392) rispetto a quella osservata per l’Italia (0.003143). Tale differenza riflette due configurazioni strutturali distinte: da un lato, una rete britannica estesa ma caratterizzata da un’elevata dispersione delle relazioni collaborative; dall’altro, un ecosistema italiano di dimensioni più contenute ma relativamente più denso e coeso.

4.2.1 Analisi multidimensionale e interpretazione dei pattern europei

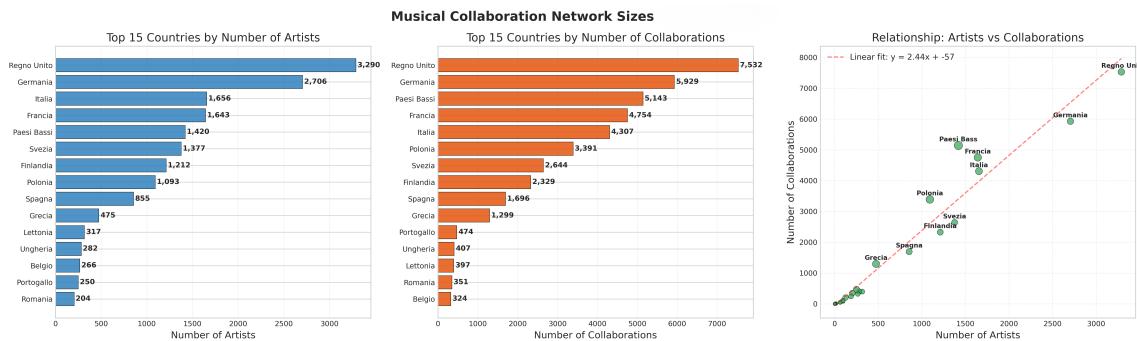


Figure 1: Analisi multidimensionale comparativa delle reti di collaborazione musicale a livello europeo.

La Figura 1 presenta un’analisi multidimensionale comparativa delle reti di collaborazione musicale a livello europeo. L’analisi sintetizza tre aspetti critici della struttura delle reti europee: dimensione, intensità collaborativa e relazione tra queste variabili. Gli istogrammi presentano i primi 15 paesi europei per due metriche distinte: a sinistra il numero di artisti, al centro il numero totale di collaborazioni. In quest’ultimo si nota:

- Il **Regno Unito** (7.532 collaborazioni) e la **Germania** (5.927) si collocano ai vertici per numero complessivo di collaborazioni; tale risultato è principalmente riconducibile alla loro elevata numerosità di nodi.
- I **Paesi Bassi** (5.143 collaborazioni) rappresentano il caso più rilevante: pur essendo quinti per numero di artisti, si posizionano terzi per volume di collaborazioni, superando sia l’Italia (4.307) che la Francia (4.754). Ciò indica un’intensa attività collaborativa.
- La **Polonia** (3.391 collaborazioni) mostra un’intensità sorprendente, superando paesi con reti più estese come Svezia (2.644) e Finlandia (2.329).

4.2.2 Relazione Strutturale tra Artisti e Collaborazioni

Il grafico a destra esplora la relazione fondamentale tra la dimensione di una rete (numero di artisti) e la sua attività (numero di collaborazioni). La relazione tra il numero di artisti e il numero totale di collaborazioni è stata analizzata mediante una regressione lineare stimata con il metodo dei minimi quadrati, considerando tutti i paesi europei inclusi nello studio. La retta ottenuta, $y = 2.44x - 76$, descrive l’andamento medio delle reti musicali europee.

- **Coefficiente angolare (2.44)**: indica che, in media, a ogni artista aggiuntivo corrispondono circa 2.44 collaborazioni in più. Questo valore esprime l’intensità collaborativa media dei network musicali analizzati.
- **Intercetta (-76)**: il valore negativo suggerisce che reti molto piccole tendono ad avere un numero di collaborazioni inferiore a quello previsto dalla tendenza generale. In particolare, al di sotto di circa 30 artisti, la struttura della rete risulta meno attiva.

La retta di regressione può quindi essere utilizzata come riferimento: i paesi posizionati al di sopra mostrano un livello di collaborazione superiore a quanto atteso rispetto alle loro dimensioni, mentre quelli al di sotto presentano un’attività collaborativa più contenuta.

- I paesi che si collocano **sopra la retta di regressione** (Paesi Bassi, Polonia, Grecia) rappresentano ecosistemi iper-collaborativi, dove il volume di interazioni supera sistematicamente le aspettative date le dimensioni. Questi sistemi sono tipicamente caratterizzati da alti valori di densità e coefficiente di clustering.
- I paesi **prossimi alla retta** (Italia, Francia, Germania, Regno Unito) seguono una relazione approssimativamente lineare, dell’attività collaborativa in funzione della dimensione.
- La **dispersione** dei dati conferma l’assenza di un modello unico europeo, evidenziando invece una pluralità di configurazioni strutturali.

4.3 Analisi comparativa con i principali paesi non-europei

Per collocare l'ecosistema musicale italiano in un contesto globale più ampio, è stata condotta un'analisi comparativa estesa a 49 paesi non-europei.

La Tabella 3 presenta il confronto tra l'Italia e i primi 5 paesi non-europei per dimensione della rete.

Table 3: Confronto con i primi 5 paesi non-europei per dimensione della rete

Metrica	Italia	Stati Uniti (1°)	Brasile (2°)	India (3°)	Giappone (4°)	Messico (5°)
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

L'analisi rivela una marcata asimmetria dimensionale tra i sistemi musicali analizzati:

- Gli **Stati Uniti** presentano caratteristiche dimensionali eccezionali, con una rete di 6.217 artisti e 14.860 collaborazioni, oltre 3,7 volte più grande della rete italiana. Questa dimensione considerevole è accompagnata da una bassa densità (0.000769) che riflette l'ampia scala e diversificazione del mercato musicale statunitense, con ben 111 gruppi distinti che operano in modo relativamente isolato. Il grado medio (4.78) è leggermente inferiore a quello italiano (5.20), indicando che, nonostante le dimensioni maggiori, la rete statunitense mostra una minore intensità di connessioni per artista.
- Il **Brasile** (1.859 artisti) e l'**India** (999 artisti) rappresentano i maggiori sistemi musicali emergenti, entrambi caratterizzati da reti ad alta densità (0.003212 e 0.005685 rispettivamente) e da un'elevata coesione strutturale. Questi valori indicano ecosistemi musicali altamente integrati, con soli 3 e 2 gruppi distinti rispettivamente, che riflettono una forte unità culturale e geografica.
- L'Italia si posiziona al sesto posto a livello globale non-europeo per dimensione, superando paesi come Corea del Sud (708) e Australia (653). Rispetto al Giappone (890), l'Italia mostra una struttura molto più coesa: il Giappone, pur avendo più artisti, presenta una rete più frammentata (38 gruppi distinti) e un grado medio inferiore (2.48), mentre l'Italia, con 1.656 artisti, mantiene una forte integrazione (16 componenti) e un'elevata attività collaborativa (grado medio 5.20).

L'analisi identifica diversi modelli di ecosistemi musicali caratterizzati da intensa attività collaborativa:

- **Porto Rico** rappresenta il caso più marcato di iper-collaboratività, con un grado medio particolarmente elevato (10.66) e un coefficiente di clustering alto (0.314).
- Il **Brasile** e l'**India** mostrano strutture simili, con gradi medi elevati (5.97 e 5.67 rispettivamente) e clustering significativo (0.173 e 0.181). Questi valori suggeriscono ecosistemi vibranti con forte tendenza alla formazione di comunità coese, supportata da una bassa frammentazione (3 e 2 gruppi distinti) che indica scene musicali unificate e ben integrate.

- La **Corea del Sud** presenta un caso interessante di rete moderatamente grande (708 artisti) ma con alta densità (0.006185) e grado medio significativo (4.37), riflettendo una scena musicale strutturata e interconnessa.
- L’Italia, con un grado medio di 5.20, si colloca tra i paesi con maggiore attività collaborativa a livello globale, superando paesi come Stati Uniti (4.78), Germania (4.38), Regno Unito (4.58) e Corea del Sud (4.37). Questo posizionamento indica una cultura musicale particolarmente orientata alla collaborazione, che compensa le dimensioni più contenute della rete con una maggiore intensità di interazioni.

4.3.1 Densità e frammentazione

- I paesi **africani** analizzati (Ghana, Egitto, Nigeria, Algeria) rappresentano il modello delle reti piccole ultra-dense, con densità che vanno da 0.047 a 0.143. Queste reti, pur avendo un numero limitato di artisti (14-66 nodi), mostrano un’elevatissima interconnessione interna. In particolare, Ghana (0.138) e Algeria (0.143) presentano densità oltre 40 volte superiori a quella statunitense (0.000769), indicando scene musicali estremamente compatte dove quasi tutti gli artisti collaborano direttamente tra loro.
- Le reti **latinoamericane** mostrano densità intermedie ma significativamente superiori alle reti europee e nordamericane: Venezuela (0.077), Panama (0.065), Colombia (0.010), Argentina (0.010). Questi valori, compresi tra 3 e 10 volte la densità italiana, riflettono scene musicali regionalmente coese ma sufficientemente ampie da sostenere una certa diversificazione interna. La bassa frammentazione (2-5 componenti connesse) conferma l’alto grado di integrazione di queste reti.
- L’Italia (densità 0.00314) si colloca in una posizione intermedia nel panorama globale. Se confrontata con i primi 5 paesi non-europei per dimensione, presenta una densità:
 - 4 volte superiore a quella degli Stati Uniti (0.000769)
 - simile a quella del Brasile (0.003212)
 - circa la metà di quella dell’India (0.005685) e del Messico (0.005580)
 - superiore a quella del Giappone (0.002788)

Questo posizionamento riflette una rete che bilancia efficacemente dimensione (1.656 nodi) e coesione.

- Tra i paesi con reti di dimensioni simili all’Italia (1.000-2.000 nodi), l’India (999 nodi, densità 0.005685) e la Corea del Sud (708 nodi, densità 0.006185) mostrano densità quasi doppie rispetto all’Italia, riflettendo modelli di collaborazione più intensi. Tuttavia, l’Italia compensa con un grado medio (5.20) superiore a quello di molti di questi paesi, indicando che pur con una densità complessiva moderata, gli artisti italiani tendono a collaborare con un numero maggiore di colleghi.

4.3.2 Clustering e coesione Locale

- I più alti valori di **clustering coefficient** si osservano in **Porto Rico** (0.314), **Ghana** (0.287) e **Repubblica Dominicana** (0.275). Questi valori eccezionalmente elevati, compresi tra 8 e 11 volte quelli di paesi come Canada o Australia, indicano una forte tendenza

alla formazione di "triadi chiuse" nelle quali i collaboratori di un artista tendono a collaborare frequentemente anche tra loro. Questo modello suggerisce comunità musicali estremamente coese.

- **India** (0.181) e **Brasile** (0.173) mostrano valori di clustering elevati che, combinati con le dimensioni significative delle loro reti (999 e 1.859 nodi rispettivamente), riflettono scene musicali che uniscono ampiezza a forte coesione interna. Questi valori, circa 1.5-2 volte superiori a quello italiano, indicano ecosistemi musicali in cui si nota la formazione di comunità artistiche strettamente interconnesse.
- L'Italia (0.119) presenta un valore di clustering moderato ma significativo nel contesto globale. Il confronto con i principali paesi non-europei rivela che:
 - Il clustering italiano è superiore a quello degli Stati Uniti (0.082) e del Giappone (0.042)
 - È inferiore a quello di India (0.181), Brasile (0.173) e Messico (0.156)
 - Si colloca in una posizione intermedia tra i paesi con reti di dimensioni simili

Questo posizionamento indica una rete che bilancia coesione locale e apertura: sufficientemente coesa da favorire la formazione di circoli artistici stabili, ma anche abbastanza aperta da permettere nuove connessioni e scambi esterni.

- I bassi valori di clustering in paesi come **Canada** (0.037), **Australia** (0.035) e **Cina** (0.027) suggeriscono strutture reticolari meno inclini alla chiusura triadica. Questi valori, circa 3-4 volte inferiori a quello italiano.

4.4 Nodes

4.4.1 Misure di centralità

Per identificare gli artisti strutturalmente più importanti nella rete di collaborazioni, sono state calcolate quattro misure di centralità. L'obiettivo è comprendere quali artisti occupano posizioni strategiche e come queste posizioni si manifestano attraverso diversi aspetti della struttura della rete.

4.4.1.1 Degree Centrality

La **degree centrality** quantifica il numero di connessioni dirette di un nodo, normalizzato per il massimo numero possibile di connessioni. Nella rete musicale analizzata, questa misura rappresenta il numero di artisti diversi con cui un determinato artista ha collaborato.

L'implementazione calcola sia il grado assoluto, che la degree centrality normalizzata. I risultati rivelano una distribuzione fortemente asimmetrica: la degree centrality media è pari a 0.0031, mentre il valore massimo raggiunge 0.0689. Tale distribuzione indica che la maggior parte degli artisti mantiene un numero limitato di collaborazioni, mentre un ristretto gruppo di nodi concentra un numero significativamente elevato di connessioni.

La Tabella 4 presenta i dieci artisti con la degree centrality più elevata:

Guè emerge come il nodo più connesso della rete con 114 collaborazioni distinte, identificandosi come hub centrale della scena musicale italiana. La presenza di Andrea Bocelli al secondo posto con 103 collaborazioni risulta particolarmente significativa: nonostante operi in

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

un genere musicale sostanzialmente diverso (classica/pop crossover), ha sviluppato una rete estesa di collaborazioni che attraversa molteplici generi musicali. La predominanza di artisti hip hop nelle prime posizioni (Clementino, Gemitaiz, Night Skinny, Don Joe, Inoki, Fabri Fibra, Emis Killa) conferma che questo genere presenta un'elevata propensione alle collaborazioni e costituisce un elemento centrale nella struttura della rete. La presenza di Ennio Morricone in decima posizione introduce un elemento di diversità generazionale e stilistica nella classifica.

4.4.1.2 Eigenvector Centrality

La **eigenvector centrality** attribuisce importanza non solo alla quantità di connessioni, ma alla loro qualità: un artista presenta alta eigenvector centrality se risulta connesso ad altri artisti che a loro volta occupano posizioni centrali nella rete. L'algoritmo converge iterativamente, assegnando a ciascun nodo un punteggio proporzionale alla somma dei punteggi dei suoi nodi adiacenti.

I risultati mostrano una concentrazione ancora più marcata rispetto alla degree centrality: il valore medio è 0.0089 mentre il massimo raggiunge 0.2573, evidenziando che un numero ristretto di artisti forma un nucleo centrale altamente coeso.

La Tabella 5 presenta i dieci artisti con la eigenvector centrality più elevata:

Table 5: Top 10 artisti per Eigenvector Centrality

Posizione	Artista	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 mantiene la posizione dominante con un valore considerevolmente superiore agli altri artisti (circa 25% in più rispetto al secondo classificato), indicando che le sue collaborazioni

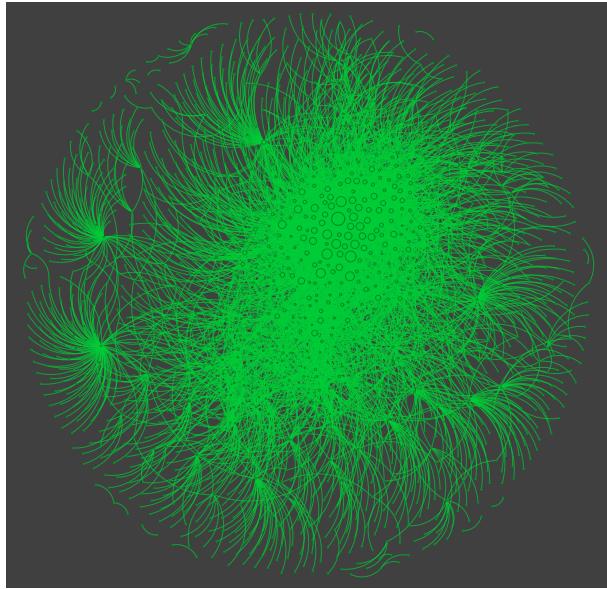


Figure 2: Caption

coinvolgono prevalentemente gli artisti più centrali della scena. La presenza esclusiva di artisti hip hop in questa classifica rivela l'esistenza di un nucleo centrale dominato da questo genere. Artisti quali Marracash, Jake La Furia, Don Joe, MadMan e Lazza formano un core altamente interconnesso che definisce il centro della rete hip hop italiana. È significativo notare che Andrea Bocelli, pur presentando un elevato numero di collaborazioni (secondo in degree centrality), risulta assente da questa classifica, suggerendo che le sue collaborazioni coinvolgono prevalentemente artisti meno centrali o più periferici rispetto al nucleo principale della rete.

4.4.1.3 Closeness Centrality

(NON MI CONVINCE) La **closeness centrality** misura la vicinanza di un nodo rispetto a tutti gli altri nodi della rete, calcolando l'inverso della distanza media basata sui cammini minimi. Un artista con elevata closeness centrality può raggiungere rapidamente qualsiasi altro artista nella rete attraverso un numero limitato di intermediari.

I risultati mostrano un valore medio di 0.2336 con un massimo di 0.3677, indicando che anche gli artisti più centrali necessitano mediamente di circa tre passaggi per raggiungere qualsiasi altro nodo della rete.

Nel contesto pratico della scena musicale, un'alta closeness centrality conferisce vantaggi strategici concreti: gli artisti ottimalmente posizionati possono accedere più facilmente a informazioni e tendenze emergenti in diverse parti della scena, avviare collaborazioni con artisti distanti nella rete attraverso pochi intermediari, e diffondere rapidamente il proprio stile o innovazioni musicali attraverso l'intero ecosistema. Questa metrica identifica quindi artisti che, indipendentemente dalla loro appartenenza al nucleo centrale, occupano posizioni che facilitano la comunicazione e la circolazione di opportunità attraverso l'intera rete.

La Tabella 6 presenta i dieci artisti con la closeness centrality più elevata:

Guè conferma la sua posizione dominante, risultando ottimalmente posizionato per accedere all'intera rete con il minimo numero di intermediari. Clementino e J-AX emergono particolarmente in questa metrica rispetto alla eigenvector centrality, suggerendo che occupano posizioni strategiche di bridging: pur non essendo nel core più denso, le loro collaborazioni attraversano diversi cluster della rete, permettendo loro di fungere da connettori efficienti tra sottocomunità

Table 6: Top 10 artisti per Closeness Centrality

Posizione	Artista	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

diverse. La presenza di Elisa (nona posizione) e Rocco Hunt (decima posizione) introduce una diversificazione di generi significativa. Questi artisti, pur non appartenendo al nucleo hip hop dominante, mantengono posizioni di vicinanza globale che facilitano la trasmissione di influenze cross-genre. La rete, sebbene dominata dall'hip hop nel nucleo centrale, mantiene quindi una struttura che permette comunicazione efficiente tra generi diversi, suggerendo un ecosistema musicale dove le barriere tra comunità di genere non impediscono la circolazione rapida di idee e opportunità collaborative.

4.4.1.4 Betweenness Centrality

La **betweenness centrality** identifica i nodi che si trovano frequentemente sui cammini minimi tra altre coppie di nodi. Artisti con elevata betweenness fungono da intermediari tra diverse componenti della rete, anche senza necessariamente presentare un numero elevato di connessioni dirette. Il calcolo è stato approssimato campionando 1000 nodi casuali per ragioni di efficienza computazionale.

I risultati mostrano una distribuzione fortemente asimmetrica: il valore medio è 0.0018 mentre il massimo raggiunge 0.1250, indicando che un numero molto limitato di nodi controlla i flussi di connessione tra cluster diversi della rete.

Andrea Bocelli domina questa metrica con un valore significativamente superiore agli altri artisti (quasi 40% superiore rispetto al secondo classificato), rivelando il suo ruolo strutturale come ponte tra contesti musicali differenti. Nonostante non appartenga al core hip hop identificato dalla eigenvector centrality, Bocelli connette la musica classica e pop con altri generi, fungendo da intermediario essenziale nella struttura complessiva della rete. La presenza di Ennio Morricone (quinta posizione), DJ Matrix (sesta posizione), Jovanotti (nona posizione) e Cristina D'Avena (decima posizione) - artisti che operano in generi diversi dall'hip hop - conferma che la betweenness centrality cattura un ruolo strutturale distinto dalla centralità locale. Questi artisti non appartengono al nucleo centrale ma occupano posizioni di gatekeeping tra comunità di genere diverse. Clementino mantiene una posizione elevata in tutte le metriche analizzate (terzo in degree, secondo in closeness e betweenness), confermandosi come uno degli artisti strutturalmente più rilevanti della rete: parte integrante del core hip hop ma con spiccate capacità di bridging verso altre scene musicali.

La Tabella 7 presenta i dieci artisti con la betweenness centrality più elevata:

Table 7: Top 10 artisti per Betweenness Centrality

Posizione	Artista	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 Assortatività

L'analisi di assortatività esamina le tendenze di connessione basate su attributi specifici dei nodi. Sono state calcolate tre forme di assortatività per caratterizzare i pattern di collaborazione nella rete musicale italiana.

4.4.2.1 Degree Assortativity

Il coefficiente di degree assortativity ottenuto è **-0.1052**, indicando una rete leggermente **disassortativa**. Questo valore negativo significa che artisti con numerose collaborazioni tendono a connettersi con artisti che ne hanno meno, piuttosto che collaborare prevalentemente tra loro. Nel contesto della rete musicale analizzata, questo pattern suggerisce che gli hub (artisti con elevato grado) non formano un gruppo isolato ma includono nelle loro collaborazioni anche artisti meno prolifici. Questo può riflettere diverse dinamiche: artisti affermati che forniscono visibilità ad artisti emergenti, producer e featuring artist che collaborano con un ampio spettro di artisti a diversi livelli di attività, o più in generale una scena musicale relativamente aperta dove il numero di collaborazioni pregresse non costituisce una barriera significativa per future opportunità. Il valore negativo, seppur moderato, è coerente con una struttura che facilita la mobilità e l'accesso anche per artisti meno centrali, in contrasto con modelli rigidamente gerarchici dove esclusivamente le figure di maggior successo collaborano tra loro.

4.4.2.2 Followers Assortativity

Il coefficiente di followers assortativity è **0.0724**, un valore positivo ma prossimo allo zero, indicando una tendenza molto debole verso pattern assortativi basati sulla popolarità.

Questo risultato suggerisce che la popolarità, misurata attraverso il numero di follower su Spotify, esercita un'influenza limitata sui pattern di collaborazione. Artisti con un elevato numero di follower non mostrano una marcata preferenza per collaborazioni esclusive con altri artisti di analogia popolarità, né emerge una segregazione netta tra diversi strati di popolarità. L'analisi dettagliata categorizza gli archi in High-High, Low-Low e High-Low utilizzando la mediana come soglia, rivelando che esistono sia collaborazioni tra artisti di analogia popolarità che collaborazioni con artisti con un grado di popolarità diverso dal proprio, in proporzioni relativamente bilanciate.

Questo pattern indica un ecosistema musicale italiano relativamente fluido rispetto al criterio

della fama: le collaborazioni non risultano fortemente vincolate dal numero di follower, e artisti affermati mostrano disponibilità a collaborare con colleghi meno conosciuti. Tuttavia, il valore leggermente positivo suggerisce l'esistenza di una debole preferenza omofila: artisti molto popolari collaborano con maggiore frequenza con altri artisti popolari, ma questa tendenza non crea barriere strutturali significative.

4.4.2.3 Genre Assortativity e Modularity

I risultati relativi al genere musicale mostrano pattern marcatamente diversi rispetto alle altre forme di assortatività. Il coefficiente di genre assortativity è **0.4778**, un valore decisamente positivo che indica una forte tendenza assortativa. La modularità è **$Q = 0.2987$** , classificabile come moderata secondo la scala interpretativa implementata (compresa tra 0.1 e 0.3).

Questi valori rivelano che il genere musicale esercita un'influenza significativa sui pattern di collaborazione: gli artisti tendono a collaborare prevalentemente con altri artisti dello stesso genere o di generi affini. Tuttavia, la modularità moderata (non elevata) indica che questa tendenza non produce compartimenti impermeabili: esistono numerose collaborazioni cross-genre che mantengono la rete globalmente interconnessa.

La Tabella 8 presenta le dieci coppie di generi più frequenti nelle collaborazioni:

Table 8: Top 10 coppie di generi nelle collaborazioni

Pos.	Genere 1	N. collab.	Genere 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 / Colonne sonore	37	(intra-genere)
10	Elettronica / Dance	34	Hip Hop / Rap

L'Hip Hop / Rap domina con 1829 collaborazioni intra-genere, confermando l'elevata densità e coesione di questa comunità musicale. Il Pop presenta 819 collaborazioni interne, dimostrando anch'esso una significativa coesione interna. La combinazione Hip Hop / Rap + Pop conta 596 archi, costituendo la contaminazione cross-genre di gran lunga più frequente. Questo rivela un ponte strutturale fondamentale tra i due generi dominanti della scena italiana. Le combinazioni Hip Hop / Rap + Indie (104 archi) e Elettronica / Dance + Pop (62 archi) rappresentano altre contaminazioni significative, sebbene con frequenza considerevolmente inferiore. Generi quali Classica/Orchestrale (72 archi interni), Elettronica/Dance (85 archi), e Soundtrack (37 archi) evidenziano comunità di dimensioni più contenute ma comunque internamente coese.

Il rapporto tra archi intra-genere e inter-genere, calcolato dalla funzione attraverso l'analisi delle sovrapposizioni tra liste di generi, mostra che le collaborazioni rimangono prevalentemente all'interno dei confini di genere, con eccezioni rilevanti.

Nel contesto della scena musicale italiana, questi risultati descrivono un quadro di specializzazione con contaminazione selettiva: i generi mantengono identità distinte e gli artisti collaborano prevalentemente entro i propri confini stilistici, ma esistono ponti consolidati tra generi

complementari, in particolare tra hip hop e pop. L'elevata densità dell'hip hop (1829 collaborazioni interne) conferma quanto emerso dall'analisi di centralità: questo genere non solo domina il nucleo centrale della rete, ma ha sviluppato un ecosistema interno estremamente ricco e interconnesso.

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.

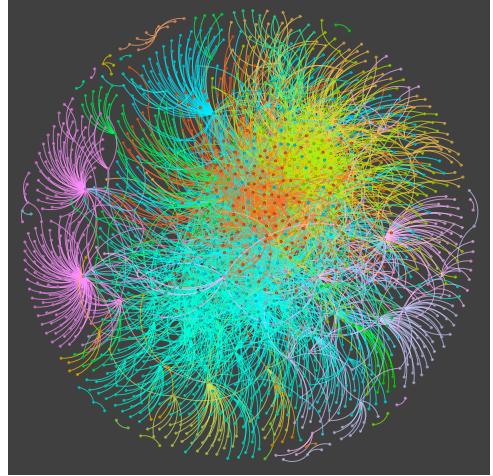


Figure 3: Community structure obtained by applying the Louvain algorithm in Gephi.

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.

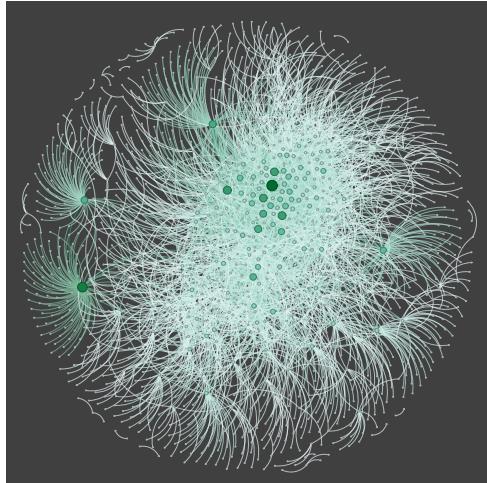
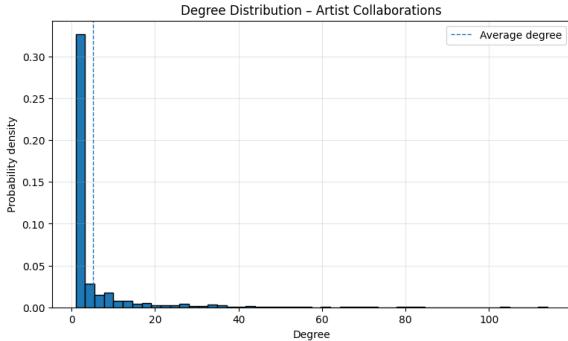
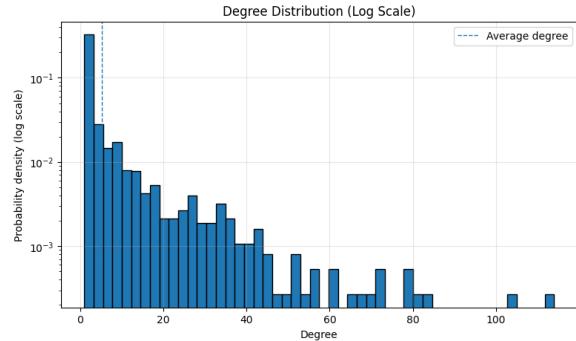


Figure 4: Artist collaboration network visualized in Gephi with node size proportional to degree.



(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

Qualitative analysis of the quantitative findings of the study.

6 Critique

Nonostante l’analisi abbia prodotto risultati significativi e interpretabili, è importante riconoscere le principali limitazioni intrinseche al nostro approccio metodologico e ai dati utilizzati.

- **Dati parziali e limitati alla piattaforma:** Il dataset include esclusivamente collaborazioni registrate ufficialmente su Spotify, escludendo quelle avvenute su altre piattaforme (YouTube, SoundCloud), in contesti live, o in forme non ufficiali. Ciò può portare a una sottostima della densità della rete, specialmente per generi underground o per artisti emergenti che operano al di fuori dei canali mainstream.
- **Inferenza automatica degli attributi:** La nazionalità e il genere musicale sono stati assegnati tramite procedure automatizzate. Sebbene questo abbia massimizzato la copertura, può aver introdotto errori sistematici o semplificazioni eccessive.
- **Analisi statica vs. dinamica:** La rete è stata trattata come un’istantanea fissa nel tempo. Un approccio longitudinale avrebbe permesso di studiare come le comunità si formano, come la centralità degli artisti cambia in risposta a eventi (uscita di un album, ingresso in classifica), e come le strategie collaborative si evolvono nelle diverse fasi della carriera.
- **Semplicità del modello di rete:** Il grafo è monomodale (solo artisti) e non orientato, e non distingue tra collaborazioni occasionali e partnership stabili. Una rappresentazione più ricca, magari pesata in base al numero di tracce condivise o arricchita con metadati temporali, avrebbe permesso analisi più accurata.
- **Limiti interpretativi dell’approccio strutturale:** L’analisi quantitativa identifica pattern di connessione, si può misurare che un artista collega due comunità musicali, ma non sappiamo se si tratti di una scelta creativa consapevole, di una relazione personale, di una strategia

commerciale. Senza integrare i dati strutturali con fonti qualitative, come interviste, analisi dei testi, ricostruzione dei contesti produttivi o dinamiche di mercato, le conclusioni sulle “strategie collaborative” o sul “ruolo sociale” degli artisti restano ipotesi plausibili ma non troppo dettagliate.

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