

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

Federico , matricola  
Federica Santisi, matricola  
Giorgia Pirelli, matricola

Dicembre 2024

## 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. Attraverso metriche di centralità, analisi di comunità e studio dei pattern di connessione, possiamo identificare quali artisti occupano posizioni strategiche, come si formano le comunità musicali e quali fattori influenzano il successo delle collaborazioni.

Questo studio si concentra sull'analisi delle collaborazioni tra artisti su Spotify utilizzando un dataset che include artisti di diverse nazionalità, con particolare attenzione, ma non esclusiva, alla scena italiana. Tuttavia, l'analisi non si limita al contesto italiano. Per comprendere appieno le dinamiche delle collaborazioni musicali è necessario adottare una prospettiva globale, esaminando come gli artisti di diverse nazionalità interagiscano tra loro, quali siano i pattern di collaborazione transnazionale e come il successo in un mercato locale possa tradursi in visibilità internazionale. In particolare, è interessante analizzare se esistano "ponti" tra scene musicali diverse, quali artisti fungano da connettori tra mercati geograficamente e culturalmente distanti, e se determinati generi musicali siano più propensi alla collaborazione internazionale rispetto ad altri.

L'obiettivo di questo studio è quindi duplice: da un lato, fornire un'analisi approfondita della struttura della rete di collaborazioni musicali su Spotify, identificando pattern, comunità e artisti chiave; dall'altro, utilizzare questa analisi per rispondere a domande di ricerca specifiche rel-

ative alla popolarità degli artisti, alle strategie di collaborazione E all'identificazione di talenti emergenti.

## 2 Problem and Motivation

L'obiettivo principale di questo studio è comprendere le dinamiche delle collaborazioni musicali e il loro impatto sul successo e sulla visibilità degli artisti. In particolare, ci proponiamo di affrontare le seguenti questioni di ricerca:

- **Identificazione del grado di popolarità degli artisti:** attraverso metriche di centralità (degree centrality, betweenness centrality, closeness centrality), si intende individuare quali artisti occupano posizioni strategiche nella rete delle collaborazioni. Un artista con alta degree centrality collabora con numerosi altri artisti, indicando una forte integrazione e un ruolo attivo nella scena musicale. Un artista con alta betweenness centrality funge da "ponte" tra diversi gruppi di artisti, potenzialmente collegando scene musicali o generi diversi e facilitando la circolazione di stili e influenze. L'obiettivo è verificare se e come queste metriche di centralità correlino con indicatori di successo commerciale quali il numero di follower su Spotify, gli stream totali e le presenze nelle classifiche globali, permettendo di comprendere se una posizione centrale nella rete di collaborazioni si traduca effettivamente in maggiore popolarità presso il pubblico.
- **Propensione alle collaborazioni transnazionali:** si intende analizzare se e in che misura gli artisti tendano a collaborare prevalentemente con artisti della stessa nazionalità o se mostrino apertura verso collaborazioni internazionali. Questo aspetto è particolarmente rilevante per comprendere le dinamiche di globalizzazione della musica contemporanea. L'analisi mira a identificare eventuali barriere linguistiche, culturali o geografiche che limitano le collaborazioni transnazionali, e a verificare se determinati generi musicali (come il rap, la musica elettronica o il pop) favoriscano una maggiore apertura internazionale rispetto ad altri. Particolare attenzione sarà dedicata all'identificazione di artisti che fungono da "ambasciatori" culturali, connettendo la propria scena nazionale con mercati esteri e facilitando lo scambio artistico tra diverse aree geografiche.
- **Identificazione di artisti emergenti tramite analisi di rete:** attraverso l'analisi della struttura della rete e l'evoluzione temporale delle metriche di centralità, si cercherà di identificare artisti emergenti, ovvero quelli che stanno rapidamente acquisendo rilevanza attraverso collaborazioni strategiche con artisti già affermati. Un artista emergente può essere caratterizzato da un pattern di crescita nelle collaborazioni con artisti di alto profilo, da un rapido incremento del numero di follower, o da una posizione nella rete che suggerisce un elevato potenziale di crescita futura. Questa analisi può fornire insights preziosi per l'industria musicale nell'identificare talenti prima che raggiungano il mainstream, permettendo a produttori e case discografiche di individuare opportunità di investimento promettenti.
- **Diffusione internazionale e successo transnazionale:** si intende esaminare il grado di penetrazione degli artisti nei mercati esteri attraverso l'analisi delle presenze nelle classifiche internazionali (chart\_hits) e la distribuzione geografica del loro pubblico. L'obiettivo è identificare quali artisti e generi musicali abbiano maggiore appeal globale e comprendere i fattori che facilitano o ostacolano il successo internazionale. In particolare, si vuole verificare se il

successo in un mercato domestico sia un prerequisito necessario per l'affermazione internazionale, o se esistano percorsi alternativi in cui artisti raggiungono popolarità all'estero prima di consolidarsi nel proprio paese d'origine. L'analisi permetterà inoltre di valutare quali generi musicali (rap/trap, pop, indie, cantautorale, elettronica) abbiano maggiore capacità di attraversare confini nazionali e in quali mercati geografici specifici ottengano maggiore successo.

- **Inferenza di generi musicali attraverso pattern di collaborazione:** per gli artisti per cui manca l'informazione sul genere musicale nel dataset, si propone di inferirlo attraverso l'analisi sistematica delle loro collaborazioni. L'approccio si basa sul presupposto che artisti che collaborano frequentemente tendano a condividere generi musicali simili o compatibili, riflettendo affinità stilistiche e artistiche. Stabilendo una soglia minima di collaborazioni e applicando tecniche di classificazione basate sulla rete (ad esempio, analizzando i generi più frequenti tra i collaboratori di un artista), sarà possibile attribuire con ragionevole confidenza un genere musicale agli artisti non classificati. Questo metodo assume che, sebbene esistano collaborazioni cross-genre, queste siano meno frequenti rispetto alle collaborazioni intra-genre, e che quindi il "vicinato" di un artista nella rete fornisca informazioni significative sulla sua identità musicale.
- **Generi musicali e propensione al successo internazionale:** si vuole identificare quali generi musicali mostrino maggiore propensione al successo sui mercati internazionali, analizzando sistematicamente la presenza nelle classifiche estere degli artisti appartenenti a ciascuna categoria di genere. L'analisi mira a comprendere se generi con forte connotazione linguistica e culturale (come il cantautorato italiano) siano strutturalmente svantaggiati rispetto a generi più "universali" come l'elettronica, il rap o il pop, o se al contrario l'autenticità e l'unicità culturale possano rappresentare un elemento distintivo che favorisce il successo in mercati di nicchia o presso specifici segmenti di pubblico. Particolare attenzione sarà dedicata all'identificazione di eventuali correlazioni tra caratteristiche del genere (presenza di testo, lingua, complessità musicale) e capacità di penetrazione internazionale.
- **Community detection e caratterizzazione per macro-generi:** applicando algoritmi di community detection (come il metodo di Louvain o il metodo di Girvan-Newman) alla rete di collaborazioni, si intende identificare gruppi di artisti densamente connessi tra loro e analizzare come questi cluster si caratterizzino rispetto ai macro-generi musicali (pop/mainstream, rap/trap/urban, rock/indie, cantautorale, elettronica). L'obiettivo è verificare se le comunità rilevate algoritmicamente corrispondano effettivamente a raggruppamenti per genere musicale, o se emergano pattern più complessi influenzati da fattori geografici, generazionali, appartenenza a specifiche etichette discografiche o affiliazioni a particolari movimenti artistici. L'analisi delle connessioni inter-comunitarie può inoltre rivelare quali generi siano più aperti alla collaborazione cross-genre, quali artisti fungano da "ponti" tra comunità diverse facilitando la contaminazione stilistica, e se esistano barriere strutturali che limitano l'interazione tra determinate scene musicali.

### 3 Datasets

The initial dataset used for this analysis was downloaded from Spotify Artist Feature & Collaboration Network [?]. 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 [kworkb.net](http://kworkb.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.

### 3.1 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`) [?] 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.

### 3.2 Aggiunta del genere artistico

## Analisi Topologica 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** (lunghezza massima del percorso geodetico) 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 ( $\langle k \rangle$ )	5.20
Densità ( $\rho$ )	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, tipica delle reti sociali su larga scala.

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. Il diametro pari a 10 conferma che anche gli artisti più distanti nella rete possono essere raggiunti attraverso un numero limitato di passaggi. Questi valori rivelano una **connettività globale efficiente**, che permette una diffusione rapida di informazioni, influenze e opportunità collaborative attraverso l'intero ecosistema musicale italiano.

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.

## Analisi Comparativa delle Reti Nazionali

Un'analisi comparativa delle sottoreti formate dagli artisti delle principali nazionalità europee rivela differenze strutturali significative, come sintetizzato nella Tabella 2.

Table 2: Metriche Comparative delle Reti di Collaborazione per Nazionalità

Parametro	Italia	Francia	Germania	Spagna	Regno Unito
Nodi ( $n$ )	1.656	1.643	2.704	855	3.290
Archi ( $m$ )	4.307	4.754	5.927	1.696	7.532
Componenti connesse	16	27	32	12	70
Grado medio	5.20	5.79	4.38	3.97	4.58
Densità ( $\rho$ )	0.00314	0.00352	0.00162	0.00465	0.00139
Coeff. di clustering medio	0.119	0.113	0.121	0.088	0.062

### 3.3 Interpretazione delle Differenze Strutturali

- **Frammentazione (Componenti Connesse):** La Spagna (12) e l'Italia (16) presentano le reti meno frammentate, suggerendo ecosistemi musicali relativamente integrati e coesi. All'estremo opposto, il Regno Unito (70) mostra una frammentazione marcata, riflesso probabile di un mercato vasto, diversificato e segmentato in numerose sottoculture e generi indipendenti. Francia (27) e Germania (32) occupano una posizione intermedia.
- **Attività Collaborativa e Compattezza:** La Francia presenta il **grado medio** più alto (5.79), seguito dall'Italia (5.20), indicando una cultura particolarmente orientata allo scambio collaborativo. La **densità**, che misura la compattezza della rete, è massima in Spagna (0.00465), segnalando una scena più compatta e potenzialmente geograficamente concentrata. La bassa densità del Regno Unito (0.00139) è coerente con la sua alta frammentazione e la dimensione del mercato.
- **Coesione Locale (Clustering):** Germania (0.121) e Italia (0.119) mostrano i valori di **clustering medio** più elevati. Ciò indica una maggiore tendenza alla formazione di triadi chiuse e comunità strette, dove i collaboratori di un artista hanno una probabilità più alta di collaborare tra loro. Al contrario, il Regno Unito (0.062) e la Spagna (0.088) presentano valori di clustering inferiori, suggerendo strutture collaborative più aperte o configurazioni maggiormente centrate su hub che connettono artisti periferici non direttamente collegati tra loro.

# Analisi Comparativa delle Reti di Collaborazione a Livello Europeo

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

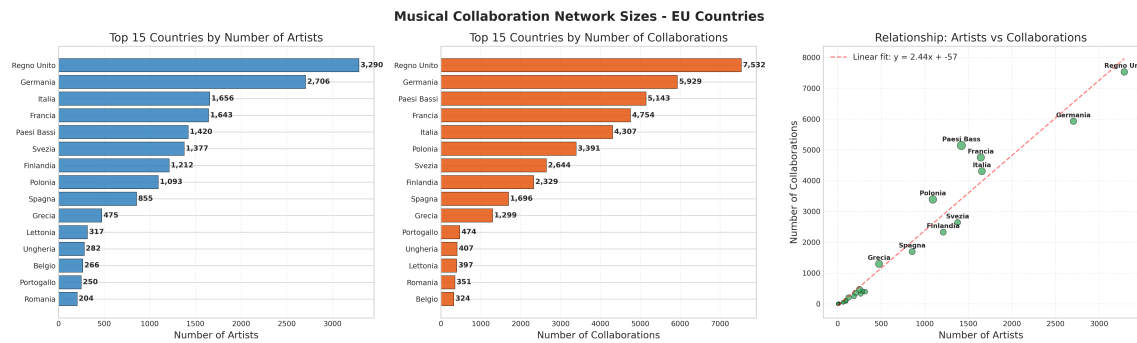


Figure 1: Analisi multidimensionale comparativa delle reti di collaborazione musicale a livello europeo.

## Panorama Dimensionale delle Reti Europee

L'istogramma a sinistra descrive la distribuzione del numero di artisti tra i paesi europei, con valori che vanno da poche centinaia a diverse migliaia.

Il **Regno Unito** emerge come il sistema di maggiori dimensioni, con 3.290 artisti, ma presenta 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.

## Intensità delle Collaborazioni

L'istogramma al centro visualizza il numero totale di collaborazioni per i primi 15 paesi.

- 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).

## 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, mostrando una scalabilità “normativa” 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 Validity and Reliability

How closely does the model of your dataset represent reality (validity)? How does the way you treat the data affect the reproducibility of the study (reliability)?

## 5 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.

## 5.1 Nodes

### 5.1.1 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 po-

sizioni strategiche e come queste posizioni si manifestano attraverso diversi aspetti della struttura della rete.

### 5.1.1.1 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 3 presenta i dieci artisti con la degree centrality più elevata:

Table 3: 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

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 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.

### 5.1.1.2 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 4 presenta i dieci artisti con la eigenvector centrality più elevata:

Table 4: 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 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.

### 5.1.1.3 Closeness Centrality

(NON MI CONVINCERE) 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 5 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 diverse. La presenza di Elisa (nona posizione) e Rocco Hunt (decima posizione) introduce

Table 5: 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

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.

#### 5.1.1.4 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 6 presenta i dieci artisti con la betweenness centrality più elevata:

Table 6: 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

### 5.1.2 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.

#### 5.1.2.1 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.

#### 5.1.2.2 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 analoga 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 analoga 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.

### 5.1.2.3 Genre Assortativity e Modularity

I risultati relativi al genere musicale mostrano pattern marcatamente diversi rispetto alle altre forme di assortatività. Il coefficiente di genere 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 7 presenta le dieci coppie di generi più frequenti nelle collaborazioni:

Table 7: 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.

## 5.2 Network

### 5.2.1 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 notable degree of genre homogeneity, particularly for the *Hip Hop / Rap* and *Pop* macro-genres, with purity values exceeding 0.6 in some cases and reaching 1.0 in smaller communities. Nevertheless, a substantial number of communities display a mixed composition, characterized by the coexistence of multiple dominant macro-genres. This suggests that artist collaborations are not strictly constrained by genre boundaries, especially within larger clusters, where cross-genre interactions appear to be more prevalent.

#### Analysis with Section 2.2.3

TODO

#### 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.

### 5.2.2 Degree Distribution

The analysis of the degree distribution of the artist collaboration network provides relevant insights into the global structure of the graph. The minimum degree of 1 indicates the presence of artists involved in a single collaboration, while the maximum degree of 114 highlights the existence of a small number of highly connected nodes acting as hubs within the network. The average degree, equal to 5.20, suggests that the network is overall sparse, with most artists collaborating with only a limited number of peers.

As shown in Figure 2, the degree distribution, examined both on a linear scale (Figure 2a) and a logarithmic scale (Figure 2b), exhibits a strongly right-skewed shape with a pronounced long tail. This behavior is characteristic of complex, scale-free or otherwise heterogeneous networks, where many low-degree nodes coexist with a few extremely high-degree nodes. In the musical context, this reflects a system in which most artists engage in occasional collaborations, while a restricted set of central artists concentrates a large number of collaborations, acting as key connectors across different regions of the network and potentially across distinct musical genres.

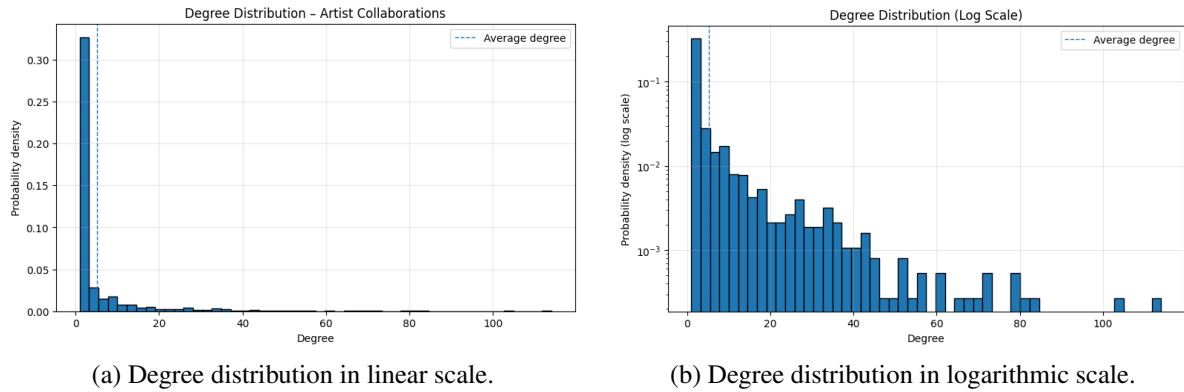


Figure 2: Degree distribution of the artist collaboration network. The linear-scale histogram highlights the strong concentration of low-degree nodes, while the logarithmic-scale representation emphasizes the long-tailed behavior caused by highly connected artists (hubs).

## 6 Conclusion

Qualitative analysis of the quantitative findings of the study.

## 7 Critique

Do you think your work solves the problem presented above? To which extent (completely, what parts)? Why? What could you have done differently to answer your research problems (e.g., gather data with additional information, build your model differently, apply alternative measures)?