

# 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

### 3.3 Analisi Topologica della Rete delle Collaborazioni Musicali Italiane

Per delineare le principali caratteristiche strutturali della rete di collaborazioni musicali italiana e fornire una panoramica completa della sua topologia, sono state calcolate una serie di misure generali di Social Network Analysis. Queste metriche permettono di comprendere le dinamiche interne della rete, rivelando pattern di connettività, efficienza nella diffusione dell'informazione e coesione tra gli artisti.

L'analisi evidenzia una rete composta da 1.656 artisti (nodi) e 4.307 collaborazioni (archi). Come mostrato nella Tabella 1, la struttura presenta diversi aspetti interessanti.

Table 1: Metriche strutturali della rete di collaborazioni musicali italiane su Spotify

Parametro	Valore
Nodi totali	1.656
Archi totali	4.307
Componenti connesse	16
Dimensione componente gigante	1.612 (97.3%)
Grado massimo	114
Grado medio	5.20
Densità	0.0031
Diametro (componente gigante)	10
Lunghezza media del percorso più breve	4.14
Coefficiente di clustering medio	0.119
Transitività	0.128
Coefficiente di assortatività	-0.105

La presenza di 16 componenti connesse indica l'esistenza di sottogruppi isolati all'interno della rete. La componente gigante, che comprende 1.612 artisti (97.3% del totale), rappresenta il nucleo principale della scena musicale italiana, mentre le altre 15 componenti più piccole corrispondono a gruppi di artisti con collaborazioni limitate o specializzate che non si connettono al resto della rete. La più grande tra queste componenti minori contiene 11 artisti, seguita da componenti di 4 e 3 nodi.

Il grado massimo di 114, attribuito all'artista Guè, indica la presenza di veri e propri *hub* nella rete—artisti che collaborano estensivamente con molti colleghi. Il grado medio di 5.20 suggerisce che ogni artista italiano collabora in media con circa 5 altri artisti, una cifra che riflette una moderata propensione alle collaborazioni nel panorama musicale nazionale.

La densità estremamente bassa (0.0031) conferma che la rete è sparsa, con solo lo 0.31% delle possibili connessioni realizzate. Questo pattern è tipico delle reti sociali su larga scala, dove il numero effettivo di connessioni è molto inferiore al potenziale massimo teorico. La distribuzione del grado mostra una struttura eterogenea: il 54.6% degli artisti (904 nodi) ha grado 1, indicando che collaborano con un solo altro artista, mentre solo il 12.6% (208 nodi) ha grado superiore a 10, rappresentando i collaboratori più attivi della scena.

Le misure di percorso rivelano una struttura relativamente efficiente. Il diametro di 10 nella componente gigante significa che la massima distanza tra due artisti è di 10 passaggi (collaborazioni), mentre la lunghezza media del percorso più breve di 4.14 indica che, in media, due artisti qualsiasi nella rete principale sono separati da poco più di 4 intermediari. Questi valori suggeriscono che informazioni, influenze musicali e opportunità di collaborazione possono

diffondersi in modo relativamente efficiente attraverso la rete.

Il coefficiente di clustering medio (0.119) e la transitività (0.128) forniscono indicazioni sulla coesione locale della rete. Entrambi i valori sono moderatamente bassi, suggerendo una limitata tendenza alla formazione di "triadi chiuse"—situazioni in cui due collaboratori di uno stesso artista tendono a collaborare anche tra loro. La transitività leggermente superiore al coefficiente di clustering medio indica la presenza di hub altamente connessi che mediano molte collaborazioni senza necessariamente formare gruppi completamente interconnessi.

Il coefficiente di assortatività negativo (-0.105) rivela una struttura di tipo "hub-spoke": gli artisti con molte collaborazioni (hub) tendono a connettersi con artisti con poche collaborazioni (spoke), piuttosto che con altri hub. Questo pattern suggerisce una certa gerarchia nella rete, dove artisti molto popolari o influenti fungono da punti focali per numerosi artisti meno connessi.

Nel complesso, l'analisi rivela una rete di collaborazioni musicali italiane caratterizzata da una struttura sparsa ma ben connessa, con una chiara gerarchia tra artisti altamente connessi (hub) e artisti con poche collaborazioni. La presenza di una componente gigante dominante indica una scena musicale sostanzialmente integrata, mentre le componenti minori suggeriscono l'esistenza di nicchie o sottocomunità specializzate. Questa struttura probabilmente facilita sia la diffusione rapida di tendenze musicali attraverso i principali hub, sia la preservazione di identità musicali distinte nelle componenti periferiche.

### 3.4 Analisi Comparativa delle Reti Nazionali di Collaborazione Musicale

L'analisi comparativa delle reti di collaborazione musicale su Spotify rivela significative differenze strutturali tra i principali paesi europei. I risultati, sintetizzati nella Tabella 2, forniscono insights sulle diverse dinamiche di collaborazione che caratterizzano le varie scene musicali nazionali.

Table 2: Metriche comparative delle reti di collaborazione musicale per paese

Parametro	Italia	Francia	Germania	Spagna	Regno Unito
Artisti totali nel dataset	2.707	3.020	4.583	1.694	6.868
Nodi nella rete	1.656	1.643	2.704	855	3.290
Archi (collaborazioni)	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	0.00314	0.00352	0.00162	0.00465	0.00139
Coeff. clustering medio	0.119	0.113	0.121	0.088	0.062

#### 3.4.1 Interpretazione dei Risultati

##### 1. Integrazione e frammentazione delle scene musicali

Il numero di componenti connesse varia notevolmente tra i paesi, riflettendo diversi gradi di integrazione delle rispettive scene musicali. La Spagna presenta la rete più integrata con solo 12 componenti, seguita dall'Italia (16), Francia (27), Germania (32) e Regno Unito (70). L'elevato numero di componenti nel Regno Unito suggerisce una scena musicale particolarmente frammentata, con numerosi sottogruppi isolati che potrebbero corrispondere a generi, sottogeneri o comunità musicali scarsamente interconnesse tra loro. Questa frammentazione potrebbe riflet-

tere la maggiore diversità culturale e musicale del mercato britannico, nonché la presenza di numerose sottoculture musicali indipendenti.

## **2. Propensione alle collaborazioni**

Il grado medio, che misura il numero medio di collaborazioni per artista, presenta valori differenziati: la Francia mostra la propensione più alta (5.79), seguita dall'Italia (5.20), Regno Unito (4.58), Germania (4.38) e Spagna (3.97). Questi dati suggeriscono che la scena musicale francese sia particolarmente orientata alle collaborazioni, mentre quella spagnola sembra più contenuta in questo aspetto. Interessante notare che, nonostante il Regno Unito abbia il maggior numero assoluto di artisti nel dataset (6.868), il grado medio rimane moderato, indicando che l'ampiezza della scena non si traduce necessariamente in una maggiore densità di collaborazioni.

## **3. Densità e sparsità delle reti**

La densità, che misura la proporzione di connessioni effettive rispetto al massimo teorico possibile, rivela strutture di rete sostanzialmente sparse in tutti i paesi, tipiche delle reti sociali su larga scala. La Spagna presenta la densità più elevata (0.00465), suggerendo una rete relativamente più coesa nonostante il minor numero assoluto di artisti. Al contrario, il Regno Unito mostra la densità più bassa (0.00139), coerente con la sua elevata frammentazione. Questa bassa densità potrebbe indicare la presenza di numerosi "mondi piccoli" all'interno della scena musicale britannica, con poche connessioni tra diverse comunità.

## **4. Coesione locale e formazione di comunità**

Il coefficiente di clustering medio, che misura la tendenza alla formazione di triadi chiuse (se A collabora con B e C, allora B e C tendono a collaborare tra loro), mostra valori moderati in tutti i paesi. La Germania presenta il coefficiente più alto (0.121), seguita da Italia (0.119), Francia (0.113), Spagna (0.088) e Regno Unito (0.062). I valori più bassi nel Regno Unito e in Spagna suggeriscono strutture di rete più "a stella", dove gli artisti collaborano con hub centrali ma non necessariamente tra loro. Questo pattern potrebbe riflettere l'influenza di grandi etichette discografiche o produttori che agiscono come connector centrali.

### **3.4.2 Considerazioni sulle Differenze Nazionali**

#### **Scena italiana: equilibrio tra integrazione e specializzazione**

La rete italiana mostra un buon equilibrio tra integrazione (16 componenti) e propensione alle collaborazioni (grado medio 5.20). La componente gigante comprende il 97.3% degli artisti, indicando una scena musicale sostanzialmente ben integrata. Il moderato coefficiente di clustering (0.119) suggerisce la presenza sia di comunità coese che di collaborazioni trasversali tra diversi gruppi musicali.

#### **Scena francese: alta propensione collaborativa**

La Francia emerge come il paese con la maggiore propensione alle collaborazioni (grado medio 5.79), nonostante una frammentazione intermedia (27 componenti). Questo potrebbe riflettere una cultura musicale particolarmente aperta alle collaborazioni cross-genre o la presenza di festival ed eventi che favoriscono incontri tra artisti di diverse background.

#### **Scena tedesca: struttura frammentata ma coesa localmente**

La Germania presenta la rete più ampia (2.704 nodi) tra i paesi europei analizzati, con un grado medio relativamente basso (4.38) ma il coefficiente di clustering più alto (0.121). Questa combinazione suggerisce la presenza di numerose comunità ben coese internamente ma con limitate connessioni reciproche, forse riflettendo la diversità linguistica e culturale regionale del paese.

#### **Scena spagnola: rete compatta e integrata**

Nonostante il minor numero assoluto di artisti (855 nodi), la Spagna mostra la rete più densa (0.00465) e meglio integrata (solo 12 componenti). Questo potrebbe indicare una scena musicale più omogenea o concentrata geograficamente, con un numero limitato di hub che connettono efficacemente l'intera rete.

#### **Scena britannica: estrema diversificazione**

Il Regno Unito presenta la scena più complessa e diversificata, con il maggior numero di artisti (3.290 nodi) ma anche la maggiore frammentazione (70 componenti) e la densità più bassa. Questa struttura riflette probabilmente la ricchezza e diversità della scena musicale britannica, con numerosi generi e sottogeneri che coesistono con limitate interazioni reciproche.

### **3.4.3 Implicazioni per l'Industria Musicale**

Queste differenze strutturali hanno importanti implicazioni per strategie di marketing, promozione e sviluppo artistico:

- **Marketing differenziato:** Le strategie di promozione dovrebbero adattarsi alle specifiche caratteristiche strutturali di ciascun mercato. In paesi come la Spagna e l'Italia, con reti più integrate, le campagne di marketing potrebbero essere più efficaci se focalizzate sugli hub centrali. Nei mercati più frammentati come il Regno Unito, potrebbe essere necessario adottare approcci più segmentati per raggiungere diverse comunità.
- **Strategie di collaborazione:** La maggiore propensione alle collaborazioni in Francia suggerisce che gli artisti potrebbero beneficiare maggiormente di strategie collaborative in quel mercato. Al contrario, in paesi con reti più frammentate, le collaborazioni potrebbero richiedere maggior sforzo per identificare i connector giusti tra comunità diverse.
- **Identificazione di talenti emergenti:** La struttura delle reti influenza come gli artisti emergenti possono ottenere visibilità. In reti più integrate come quella spagnola, un artista emergente potrebbe ottenere visibilità più rapidamente attraverso pochi collegamenti strategici. In reti più frammentate, potrebbe essere necessario costruire presenza in più comunità simultaneamente.
- **Diffusione di tendenze musicali:** La velocità e il pattern di diffusione di nuove tendenze musicali variano in base alla struttura della rete. Nelle reti più dense e integrate, le tendenze potrebbero diffondersi più rapidamente ma anche omogeneizzarsi più facilmente. Nelle reti più frammentate, potrebbero emergere tendenze diverse in comunità separate.

In conclusione, l'analisi comparativa rivela che le scene musicali nazionali su Spotify presentano strutture di rete significativamente diverse, che riflettono probabilmente differenze culturali, storiche e organizzative nei rispettivi ecosistemi musicali. Queste differenze strutturali hanno implicazioni importanti per come la musica viene creata, distribuita e consumata in ciascun mercato.

## **3.5 Analisi delle Collaborazioni per Nazionalità**

L'analisi completa delle nazionalità presenti nel dataset Spotify rivela una distribuzione fortemente concentrata, con un numero limitato di paesi che dominano la scena musicale internazionale. I risultati evidenziano significative differenze strutturali nelle reti di collaborazione delle varie nazionalità.



### 3.5.1 Distribuzione degli Artisti per Nazionalità

Il dataset comprende 192 nazionalità uniche, ma la distribuzione è estremamente diseguale. Come mostrato nella Tabella 3, i primi 10 paesi rappresentano collettivamente il 79.5% degli artisti con nazionalità nota.

Table 3: Top 10 nazionalità per numero di artisti nel dataset Spotify

Nazionalità	Numero artisti	Percentuale
Stati Uniti	11.903	7.6%
Regno Unito	6.868	4.4%
Germania	4.583	2.9%
Francia	3.020	1.9%
Italia	2.707	1.7%
Brasile	2.413	1.5%
Svezia	2.289	1.5%
Paesi Bassi	2.276	1.5%
Giappone	2.220	1.4%
Finlandia	1.726	1.1%
<b>Totale top 10</b>	<b>37.009</b>	<b>23.6%</b>

### 3.5.2 Pattern di Collaborazione Internazionale

L'analisi delle collaborazioni su un campione di 100.000 interazioni rivela che il 21% delle collaborazioni coinvolge artisti di nazionalità diversa, il 28% coinvolge artisti della stessa nazionalità, mentre il restante 51% non può essere classificato per mancanza di informazioni sulla nazionalità di uno o entrambi gli artisti. Questo dato evidenzia un significativo grado di internazionalizzazione nella produzione musicale contemporanea.

Le collaborazioni internazionali mostrano pattern ben definiti, con gli Stati Uniti che emergono come il principale hub di collaborazioni transnazionali (Tabella 4).

Table 4: Principali coppie di nazionalità per frequenza di collaborazioni internazionali

Nazione 1	Nazione 2	Collaborazioni	Percentuale
Regno Unito	Stati Uniti	884	1.77%
Canada	Stati Uniti	283	0.57%
Germania	Stati Uniti	254	0.51%
Paesi Bassi	Stati Uniti	248	0.50%
Paesi Bassi	Regno Unito	224	0.45%
Germania	Regno Unito	200	0.40%
Francia	Stati Uniti	182	0.36%
Francia	Regno Unito	182	0.36%
Australia	Stati Uniti	178	0.36%
Australia	Regno Unito	137	0.27%

### 3.5.3 Caratteristiche Strutturali delle Reti Nazionali

L'analisi delle reti nazionali complete rivela differenze significative nella struttura delle collaborazioni interne di ciascun paese (Tabella 5).

Table 5: Caratteristiche strutturali delle reti di collaborazione nazionali

Nazionalità	Nodi	Archi	Comp.	Densità	Grado medio	Grado max	Clustering
Stati Uniti	6.217	14.860	111	0.000769	4.78	272	0.082
Regno Unito	3.290	7.532	70	0.001392	4.58	104	0.062
Germania	2.704	5.927	32	0.001622	4.38	288	0.121
Francia	1.643	4.754	27	0.003524	5.79	109	0.113
Italia	2.664	4.307	16	0.003143	5.20	114	0.119

**Analisi dei Risultati per Nazionalità Stati Uniti: Il principale hub globale** La rete statunitense è la più estesa (6.217 nodi) e presenta il maggior numero di collaborazioni interne (14.860). Nonostante la bassa densità (0.000769), tipica delle reti molto ampie, mostra una buona integrazione con il 94.4% degli artisti appartenenti alla componente gigante. Il grado medio di 4.78 indica una moderata propensione alle collaborazioni, mentre il grado massimo di 272 suggerisce la presenza di veri e propri super-hub nella scena musicale americana.

**Regno Unito: Rete frammentata ma connessa** La rete britannica (3.290 nodi, 7.532 archi) presenta un elevato numero di componenti connesse (70), indicando una scena musicale diversificata con numerose sottocomunità. La densità relativamente bassa (0.001392) e il coefficiente di clustering modesto (0.062) suggeriscono una struttura più "a stella" rispetto ad altre reti nazionali.

**Germania: Coesione locale elevata** La rete tedesca si distingue per il più alto coefficiente di clustering (0.121), indicando una forte tendenza alla formazione di comunità coese internamente. Nonostante il grado medio relativamente basso (4.38), la rete mostra una buona integrazione con il 97% degli artisti nella componente gigante.

**Francia: Alta propensione collaborativa** La Francia presenta il grado medio più alto (5.79) tra le principali nazionalità analizzate, suggerendo una cultura musicale particolarmente orientata alle collaborazioni. La densità relativamente alta (0.003524) e il buon coefficiente di clustering (0.113) indicano una rete ben coesa e integrata.

**Italia: Rete compatta e ben integrata** La rete italiana, pur essendo di dimensioni moderate (2.664 nodi), mostra una delle strutture più integrate con solo 16 componenti connesse. La percentuale di collaborazioni interne più elevata (39.2%) rispetto ad altre nazionalità suggerisce una certa autosufficienza della scena musicale italiana, sebbene non manchino significative collaborazioni internazionali.

### 3.5.4 Propensione alle Collaborazioni Internazionali

L'analisi della percentuale di collaborazioni interne rispetto al totale delle collaborazioni di ciascuna nazionalità fornisce ulteriori insights (Tabella 6).

L'Italia emerge come il paese con la maggiore propensione alle collaborazioni interne (39.2%), seguito da Francia (31.1%) e Brasile (30.7%). Questo potrebbe riflettere fattori linguistici, culturali o di mercato che favoriscono le collaborazioni nazionali rispetto a quelle internazionali. Al contrario, Regno Unito (22.3%) e Giappone (22.0%) mostrano una maggiore apertura verso collaborazioni internazionali.

### 3.5.5 Implicazioni per l'Ecosistema Musicale Globale

I risultati evidenziano un ecosistema musicale globalizzato ma con marcate differenze regionali:

Table 6: Propensione alle collaborazioni interne delle principali nazionalità

Nazionalità	Collab. interne	Collab. totali	% Interne
Italia	4.307	10.999	39.2%
Francia	4.754	15.291	31.1%
Brasile	5.547	18.050	30.7%
Paesi Bassi	5.143	16.629	30.9%
Stati Uniti	14.860	53.083	28.0%
Germania	5.927	21.095	28.1%
Svezia	2.644	9.412	28.1%
Giappone	1.103	5.004	22.0%
Regno Unito	7.532	33.713	22.3%

- **Dominanza anglo-americana:** Stati Uniti e Regno Unito rappresentano i principali hub di collaborazione internazionale, funzionando da ponti tra diverse scene musicali nazionali.
- **Diversità strutturale:** Le reti nazionali presentano strutture diverse che riflettono probabilmente differenze culturali, linguistiche e organizzative nei rispettivi mercati musicali.
- **Bilinguismo musicale:** Paesi come Francia e Italia mostrano una significativa percentuale di collaborazioni interne, suggerendo l'esistenza di ecosistemi musicali relativamente autosufficienti, sebbene integrati nella rete globale.
- **Ruolo dei piccoli paesi:** Nazioni come Paesi Bassi e Svezia, pur avendo un numero relativamente basso di artisti, mostrano un'elevata connettività internazionale, probabilmente dovuta alla maggiore propensione all'uso della lingua inglese e all'integrazione nei circuiti musicali internazionali.

Questa analisi dimostra come le caratteristiche strutturali delle reti di collaborazione riflettano complesse dinamiche culturali, linguistiche e di mercato che influenzano la produzione e la diffusione della musica contemporanea a livello globale.

## 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 Network

### 5.1.1 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.

### 5.1.2 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 1, the degree distribution, examined both on a linear scale (Figure 1a) and a logarithmic scale (Figure 1b), 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.

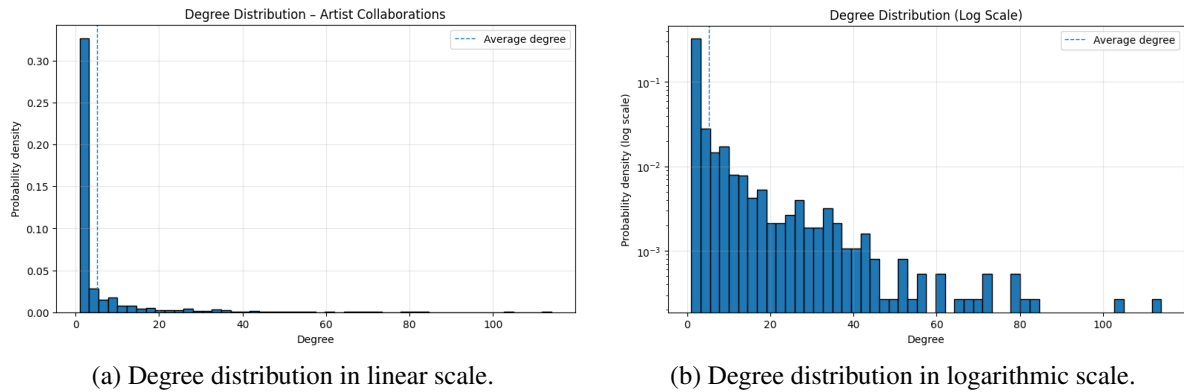


Figure 1: Degree distribution of the artist collaboration network. The linear-scale histogram highlights the strong concentration of low-degree nodes, while the logarithmic-scale representation emphasizes the long-tailed behavior caused by highly connected artists (hubs).

## 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)?