

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

Il dataset di partenza utilizzato per questa analisi è stato scaricato da Spotify Artist Feature & Collaboration Network [?]. Dopo un'attenta verifica, è stato considerato una fonte affidabile

poiché derivato da dati pubblici provenienti da Spotify API, già pre-processati per rappresentare le relazioni di collaborazione tra artisti.

Il dataset è strutturato sotto forma di grafo orientato, composto da due file principali:

- **nodes.csv** – contiene i nodi del grafo, dove ogni nodo rappresenta un artista. Le colonne includono:
 - **id**: identificativo univoco dell’artista.
 - **name**: nome dell’artista.
 - **followers**: numero di follower dell’artista su Spotify.
 - **popularity**: indice di popolarità (0–100), calcolato da Spotify in base a stream recenti e visibilità.
 - **genres**: lista dei generi associati all’artista.
- **edges.csv** – contiene gli archi del grafo, che rappresentano una collaborazione tra due artisti. Le colonne includono:
 - **source**: ID dell’artista che collabora.
 - **target**: ID dell’artista con cui collabora.
 - **weight**: valore numerico che rappresenta l’intensità della collaborazione (ad esempio numero di brani condivisi).

L’obiettivo iniziale è stato quello di arricchire il grafo con ulteriori informazioni sugli artisti, in particolare la **nazionalità** e il **genere musicale prevalente**, per poter effettuare analisi sociali e culturali più approfondite sulla rete.

3.1 Aggiunta della nazionalità dell’artista

Per associare una nazionalità agli artisti, sono state seguite due strategie complementari:

1. **Inferenza tramite genere musicale.** In un primo approccio, la nazionalità è stata dedotta analizzando il genere di appartenenza. Ad esempio, un artista associato al genere *italian hip hop* è stato etichettato come *italiano*. Questo metodo ha permesso di assegnare automaticamente una nazionalità a un sottoinsieme consistente di artisti, ma non era applicabile a tutti i casi, poiché molti generi non contenevano riferimenti geografici espliciti.
2. **Completamento tramite dataset esterno (MusicBrainz).** Per gli artisti privi di nazionalità nel primo passaggio, sono stati integrati i dati provenienti dal database MusicBrainz Dump (`mbdump`) [?]. Tuttavia, un matching diretto basato solo sul nome dell’artista presentava criticità significative (esistenza di omonimi con nazionalità diverse). Per ridurre questi errori, l’integrazione è stata effettuata solo sugli artisti rimasti non classificati dopo il primo metodo di inferenza, migliorando la precisione complessiva e preservando la coerenza dei dati.

Questa procedura ibrida ha consentito di aumentare la copertura dei dati di nazionalità e di rendere la successiva analisi più accurata, in particolare nel confronto tra comunità artistiche di diversi Paesi.

3.2 Aggiunta del genere artistico

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 Analisi sulla rete

5.1.1 Community detection

L'analisi di community detection è stata utilizzata per verificare se gli artisti tendono a collaborare prevalentemente con altri artisti appartenenti allo stesso macro-genere musicale. A tal fine sono stati applicati due approcci distinti, Louvain ed Edge Betweenness, e per ciascuno sono state valutate la numerosità delle community individuate e la loro omogeneità di genere tramite la purezza del macro-genere dominante.

Louvain

L'algoritmo di Louvain ha individuato un totale di 34 community, mostrando una struttura piuttosto frammentata della rete. Diverse community risultano caratterizzate da una discreta omogeneità di genere, in particolare per *Hip Hop / Rap* e *Pop*, con valori di purezza che in alcuni casi superano 0.6 e raggiungono 1.0 nelle community più piccole. Tuttavia, molte community presentano una composizione mista, con coesistenza di più macro-generi principali, suggerendo che le collaborazioni tra artisti non sono rigidamente vincolate al genere musicale, soprattutto nei cluster di dimensioni maggiori.

Analisi con 2.2.3

TODO

Edge Betweenness

L'algoritmo di Edge Betweenness (Girvan–Newman) ha prodotto 17 community, quindi una partizione più grossolana rispetto a Louvain. Le community ottenute risultano generalmente meno pure, con valori di purezza spesso inferiori a 0.5, specialmente nei cluster di grandi dimensioni dominati da *Pop* e *Hip Hop / Rap*. Questo indica che l'approccio basato sulla rimozione iterativa degli archi più centrali tende ad aggregare artisti di generi diversi, evidenziando ruoli di ponte e collaborazioni inter-genere piuttosto che una netta separazione per macro-genere.

5.1.2 Degree distribution

L'analisi della distribuzione dei gradi della rete di collaborazioni tra artisti fornisce indicazioni rilevanti sulla struttura globale del grafo. Il grado minimo pari a 1 indica la presenza di artisti con una singola collaborazione, mentre il grado massimo di 114 evidenzia pochi nodi altamente connessi che agiscono come veri e propri hub della rete. Il grado medio, pari a 5.20, suggerisce una rete complessivamente sparsa, in cui la maggior parte degli artisti collabora con un numero limitato di altri nodi.

La distribuzione dei gradi, osservata sia in scala lineare sia in scala logaritmica, mostra un andamento fortemente asimmetrico, con una lunga coda destra. Questo comportamento è tipico delle reti complesse di tipo scale-free o comunque eterogenee, dove coesistono molti nodi a basso grado e pochi nodi ad altissimo grado. In ambito musicale, ciò riflette un sistema in cui la maggior parte degli artisti realizza collaborazioni occasionali, mentre un numero ristretto di artisti centrali concentra un'elevata quantità di collaborazioni, fungendo da punti di connessione tra diverse aree della rete e, potenzialmente, tra generi musicali differenti.

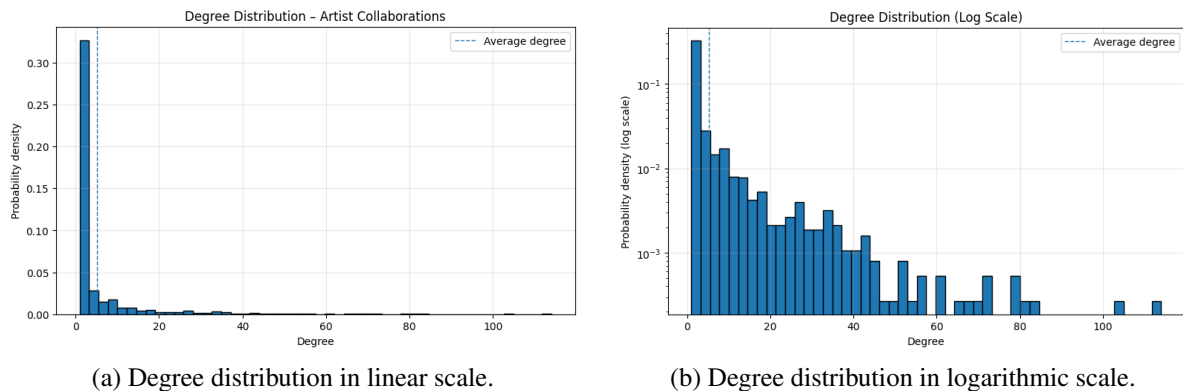


Figure 1: Degree distribution of the artist collaboration network. The linear-scale histogram highlights the strong concentration of low-degree nodes, while the logarithmic-scale representation emphasizes the long-tailed behavior caused by highly connected artists (hubs).

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)?