

Collaborazioni tra Artisti in base al Genere Musicale

Luca Boffo - 889058

Indice

1	Abstract	3
2	Introduzione	4
2.1	Contesto del Problema	4
2.2	Obiettivo del Progetto	4
2.3	Motivazione e Rilevanza della Domanda di Ricerca	4
3	Metodologia	5
3.1	Raccolta ed Analisi dei Dati	5
3.2	Elaborazioni dei Dati	5
3.3	Misure di Centralità	6
4	Processo di Analisi	8
4.1	Prima Fase	8
4.2	Seconda Fase	9
4.3	Terza Fase	12
4.4	Risultati	13
5	Limiti e Bias	14
5.1	Limiti dei Dati Utilizzati	14
6	Bibliografia	15
6.1	Dataset e Fonti di Dati	15
6.2	Software e Librerie	15
6.3	Metodologia e Concetti Teorici	15
7	Conclusioni	16

Elenco delle figure

1	Top 20 artisti in base al grado di centralità	7
2	Top 20 artisti in base alla centralità di betweenness	7
3	Collaborazioni tra artisti	8
4	Collaborazioni tra artisti con generi musicali associati	10
5	Collaborazioni tra artisti con generi musicali associati appartenenti ai top 10 generi musicali	11
6	Componente più grande tra artisti con generi musicali associati	12

1 Abstract

L'analisi che ho effettuato, si basa su una semplice domanda: "Qual è il genere musicale che si presta ad avere più collaborazioni tra artisti?"

I dati sono stati raccolti da un dataset di Spotify, al suo interno sono presenti un grande numero di tracce musicali. Assegnato ad ogni brano sono presenti caratteristiche ricercate come genere musicale a cui appartiene e il nome dell'artista. Ma anche la lunghezza in secondi del brano, il BPM (Battiti Per Minuto) e molte altre.

Dalla ricerca si denota che il genere musicale "show-tunes" è il genere che si presta ad avere più collaborazioni tra artisti, ma si nota anche che c'è una tendenza a collaborare con piccoli gruppi di artisti. Per questo motivo è presente una grande quantità di componenti del grafo.

Inoltre l'analisi dimostra la tendenza a collaborare con artisti che producono tracce musicali con generi in comune.

2 Introduzione

2.1 Contesto del Problema

L'analisi viene svolta su un dataset contenente informazioni riguardanti dei brani musicali caricati sulla piattaforma Spotify.

Il dataset comprende 114.000 tracce musicali.

Il dataset fornisce numerose informazioni, ma quelle necessarie per l'analisi sono:

- **track_genre**, il genere a cui appartiene la traccia
- **artists**, il nome dell'artista, in caso di più artisti, ognuno è separato da “;”

Altre informazioni non utili allo studio sono ad esempio `track_name`, che indica il nome del brano, `album_name`, che indica il nome dell'album a cui appartiene il brano, `duration_ms`, la durata della traccia in millisecondi e molte altre.

2.2 Obiettivo del Progetto

L'obiettivo del progetto è rispondere ad una domanda specifica: **“Qual è il genere musicale che si presta ad avere più collaborazioni tra artisti?”**.

Inoltre questo progetto ha l'intento di produrre una soluzione sotto forma di grafo. L'intento è quello di facilitare la visualizzazione della risposta al problema.

2.3 Motivazione e Rilevanza della Domanda di Ricerca

Una prima motivazione è quella di voler comprendere le dinamiche all'interno del mercato musicale. Grazie all'analisi svolta è possibile capire quali siano le mode e le tendenze nelle collaborazioni tra artisti.

Inoltre è possibile comprendere come possono evolvere i generi musicali e quali siano più in sintonia.

3 Metodologia

3.1 Raccolta ed Analisi dei Dati

Come detto in precedenza, il dataset utilizzato è composto da numerose informazioni. Il primo passaggio svolto è stata l'estrapolazione degli artisti.

Siccome l'analisi si concentra sulle collaborazioni, ho suddiviso gli artisti in due categorie, gli artisti singoli e le collaborazioni.

Ho fatto questa suddivisione, grazie ad una caratteristica del dataset, gli artisti che hanno svolto una collaborazione sono scritti in un'unica istanza, suddivisi da un ";".

Ho estrapolato tutti gli artisti in cui era presente un ";" e in seguito li ho divisi.

Successivamente ho ricavato dal mio dataset i generi delle tracce musicali, raggruppati per artisti.

3.2 Elaborazioni dei Dati

Ho effettuato i seguenti passaggi per elaborare i dati.

Una volta preso unicamente gli artisti che hanno fatto delle collaborazioni, creo un dataframe nel quale segno i legami tra due artisti.

Ad esempio, se ho un'istanza simile a "artista1;artista2", verrà divisa nel dataframe in modo tale da avere la relazione "from artista1 to artista2".

Allo stesso modo, se ho un'istanza simile a "artista1;artista2;artista3", verrà divisa nel dataframe creando tutte le varie relazioni possibili.

In seguito, conto quante volte ogni coppia ha collaborato.

3.3 Misure di Centralità

Ho eseguito un'analisi sulle misure di centralità, concentrandomi sul **Grado di Centralità** e sulla **Betwenness**.

L'analisi della centralità è stata svolta sulla componente più grande, in modo tale da mantenere un grafo completamente collegato.

Il grado di centralità è il numero di archi che un determinato nodo contiene.

Nell'ambito dell'analisi, il grado di centralità sta a significare la popolarità diretta, ovvero con quanti artisti collabora un determinato artista.

La centralità di betwenness di un nodo, misura la frequenza con cui il nodo stesso si trova sul percorso più breve tra altri due nodi del grafo.

Nell'ambito dell'analisi, identifica i nodi ponte all'interno del grafo. Un nodo ponte è un nodo che collega diverse comunità all'interno del grafo.

Come possiamo notare nel grafico a barre della figura 1, c'è una grande predominanza di artisti appartenente al genere "funk". Mentre il genere più frequente, ovvero "show-tune", ha pochi nodi con un alto grado di centralità. Infatti i generi che rappresentano gli artisti con un alto grado di centralità sono il "classical" e "opera".

La figura 2 raffigura il grafico a barre che classifica i 20 artisti con più alto fattore di betwenness. Possiamo notare che il genere musicale "show-tune", il genere musicale che si presta ad avere più collaborazioni tra artisti, ha artisti con un alto fattore di betwennes.

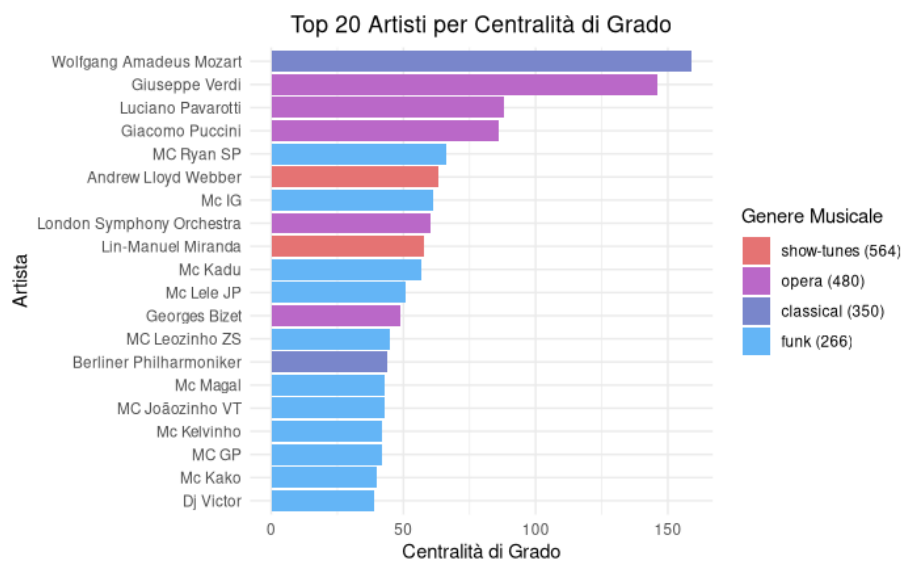


Figura 1: Top 20 artisti in base al grado di centralità

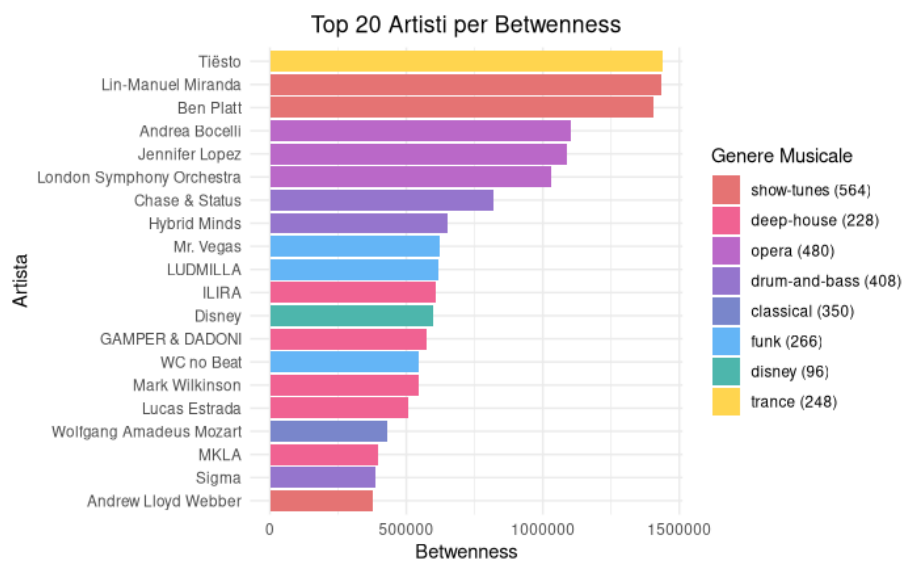


Figura 2: Top 20 artisti in base alla centralità di betweenness

4 Processo di Analisi

4.1 Prima Fase

La prima fase dell'analisi, consiste nella creazione di un grafo generale. Il quale indica le collaborazioni tra artisti.

Il grafo in questione risulta confusionario, non si delineano quali siano le singole collaborazioni.

Questo problema è dovuto all'enorme quantità di dati, anche se filtrati in precedenza.

Mantenendo unicamente gli artisti che hanno collaborato con un altro artista.

Come possiamo vedere dal grafo 3, non si riesce a ricavare molto da questa analisi.

Rete di Collaborazioni tra Artisti

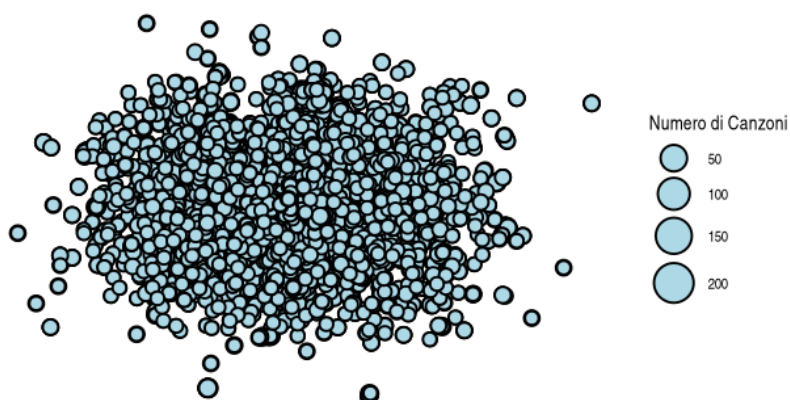


Figura 3: Collaborazioni tra artisti

4.2 Seconda Fase

La seconda fase si incentra maggiormente sull'applicazione dei generi musicali ai nodi del grafo.

Per fare ciò, ho applicato dei colori ai nodi, in modo da rendere maggiormente visibile questa informazione.

Il dataset è composto da 114 generi musicali differenti, visualizzarli tutti renderebbe meno efficiente la raffigurazione del grafo. Per questo motivo ho mantenuto i 10 generi musicali differenti più comuni. Etichettando i restanti nodi del grafo con la targa "Other".

In seguito per una migliore comprensione, ho creato una leggenda, nella quale assegno il colore al genere corrispondente e conto il numero di artisti appartenenti a quel determinato genere.

Come possiamo vedere dalla figura 4, c'è una prevalenza di generi musicali categorizzati come "other".

Questo è dovuto alla somma dei restanti nodi, non appartenente ai 10 generi musicali più comuni.

Per questo motivo, per svolgere un'analisi sui 10 generi musicali con più collaborazioni, è utile eliminare all'interno del grafo tutti i nodi appartenenti alla categoria "other". Nonostante l'approccio, possiamo notare nella figura 5 comunque una discreta confusione.

Rete di Collaborazioni tra Artisti

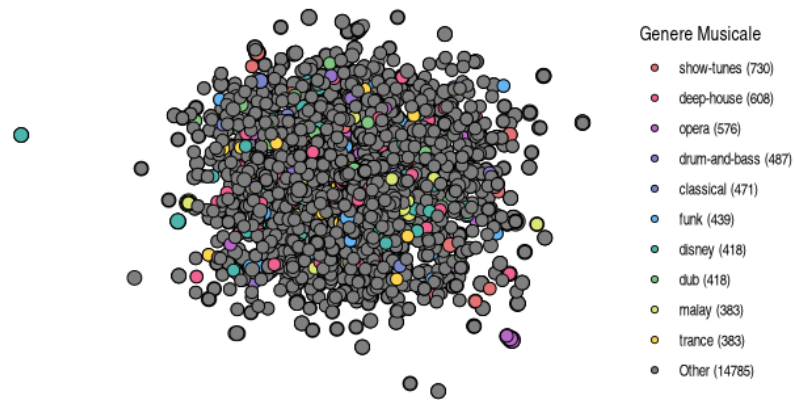


Figura 4: Collaborazioni tra artisti con generi musicali associati

Rete di Collaborazioni tra Artisti (Top 10 generi)

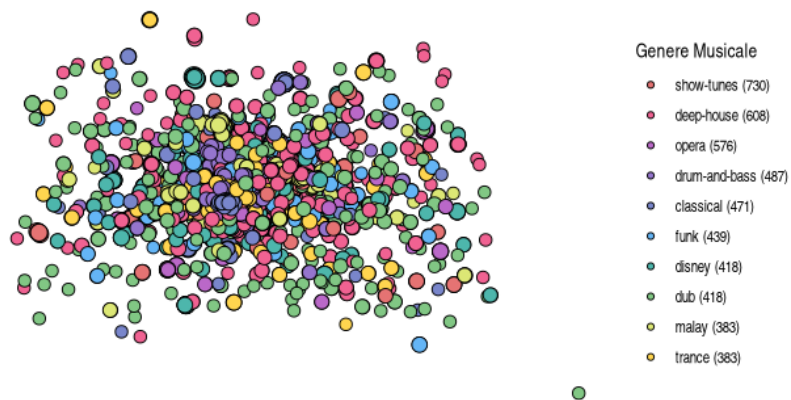


Figura 5: Collaborazioni tra artisti con generi musicali associati appartenenti ai top 10 generi musicali

4.3 Terza Fase

Per arrivare ad un risultato, è ottimale analizzare le varie componenti del grafo, ovvero i sottografi che non sono connessi tra di loro e che vanno a formare il grafo completo. Nello specifico ho analizzato la componente più grande, composta da 12267 nodi in un totale di 19698 nodi.

La componente più grande comprende circa il 62% degli artisti.

Componente più grande (Top 10 generi)



Figura 6: Componente più grande tra artisti con generi musicali associati

4.4 Risultati

Grazie alla figura 6 possiamo notare che nonostante l'utilizzo del solo 54% degli artisti, il genere musicale dominante rimane "show-tunes".

Alcuni generi musicali sono stati molto ridotti, come nel caso del genere "disney" o "dub".

Il grafo è **densamente connesso**, questo significa che gli artisti collaborano molto spesso tra di loro.

Inoltre possiamo distinguere che alcuni generi musicali tendono ad isolarsi, mentre altri collaborano con generi musicali differenti.

Un esempio di genere musicale che tende ad isolarsi è il genere "trance", a cui è associato il colore giallo e "funk", al quale è assegnato il colore celeste. Mentre il genere musicale "classical", al quale è stato assegnato il blu scuro, ha la tendenza a collaborare con generi musicali differenti. Per questo motivo i nodi associati ad esso sono sparsi. Anche il genere "drum-and-bass" è sparso all'interno del grafo, possiamo categorizzare questi generi come **ponti**. Sono generi musicali che grazie alla tendenza di collaborare con cluster differenti, formano dei collegamenti che mantengono il grafo.

Nonostante il pattern che si è venuto a creare, otteniamo come risultato che il genere musicale che si predispone ad avere più collaborazioni tra artisti è lo **show-tune**.

Lo show-tune è il genere musicale associato a canzoni scritte come colonne sonore di film e opere teatrali.

5 Limiti e Bias

5.1 Limiti dei Dati Utilizzati

L'analisi è soggetta a numerosi limiti, dovuti principalmente al dataset utilizzato.

I dati sono ristretti alla piattaforma social di Spotify.

Spotify tende ad etichettare maggiormente le tracce e i generi musicali più popolari. Questo può portare ad avere come generi più centrali i generi musicali più diffusi, rispetto a sottogeneri definiti "di nicchia".

Un'artista può appartenere a più generi musicali, per semplificare l'analisi ho assegnato ad ogni artista un unico genere musicale. Ho selezionato il genere con il maggior numero di tracce composte dall'artista appartenenti ad esso.

Il dataset fornisce i dati di tracce musicali inserite all'interno della piattaforma social dal 2000 al 2020.

L'analisi non implementa dati moderni, quindi tracce musicali inseriti all'interno di Spotify negli ultimi cinque anni.

Analizzando la componente principale, si rischia di escludere gli artisti appartenenti a generi musicali meno popolari.

6 Bibliografia

6.1 Dataset e Fonti di Dati

- [1] Spotify. *Spotify Tracks Dataset*. Visualizzato: 2025-05-16. URL: <https://www.kaggle.com/datasets/maharshipandya/-spotify-tracks-dataset>.

6.2 Software e Librerie

- [2] RStudio Team. *RStudio: Integrated Development Environment for R*. URL: <https://www.rstudio.com/>.
- [3] Gabor Csardi e Tamas Nepusz. *igraph: Network Analysis and Visualization*. URL: <https://igraph.org>.
- [4] Thomas Lin Pedersen. *ggraph: An Implementation of Grammar of Graphics for Graphs and Networks*. URL: <https://github.com/thomasp85/ggraph>.
- [5] Thomas Lin Pedersen. *tidygraph: A Tidy API for Graph Manipulation*. URL: <https://github.com/thomasp85/tidygraph>.
- [6] Hadley Wickham e Lionel Henry. *tidyr: Tidy Messy Data*. URL: <https://tidyr.tidyverse.org>.

6.3 Metodologia e Concetti Teorici

- [7] Zollo Fabiana. "Social Network Analysis - CT0540, Università Ca' Foscari di Venezia". In: 2024. URL: <https://www.unive.it/data/insegnamento/402009>.

7 Conclusioni

Questo progetto ha analizzato le dinamiche collaborative tra artisti su Spotify dal 2000 al 2020, cercando di rispondere alla domanda: "Qual è il genere musicale che si presta ad avere più collaborazioni?". Attraverso la costruzione di un grafo di rete, abbiamo esplorato la struttura delle collaborazioni.

L'analisi ha rivelato una rete con una componente principale molto ampia, includendo circa il 62% degli artisti, indicando un'elevata propensione alla collaborazione. Assegnando ad ogni artista il suo genere più prevalente, ho identificato "show-tunes" come il genere dominante nella componente principale, suggerendo una forte tendenza alla collaborazione.

Ho osservato che generi come "trance" e "funk" tendono a formare cluster più isolati nel grafo generale, sebbene singoli artisti "funk" possano avere un alto grado di centralità. Al contrario, "classical" e "drum-and-bass" emergono come "nodi ponte", cruciali per connettere diverse comunità di generi, evidenziando la loro capacità di facilitare collaborazioni inter-genere e mantenere la coesione della rete. Le misure di centralità hanno fornito ulteriori dettagli: la centralità di grado ha identificato artisti altamente connessi, in particolare nel "classical" e nell'"opera". La centralità di betweenness ha rivelato che artisti chiave nel genere "show-tunes" fungono da intermediari vitali per il flusso di interazioni, rafforzando la sua posizione centrale nelle dinamiche collaborative.

Questi risultati offrono intuizioni significative sul mercato musicale. La centralità degli "show-tunes" nelle collaborazioni suggerisce l'influenza delle produzioni cinematografiche e teatrali come principali interazioni fra diversi generi. L'esistenza di generi "ponte" evidenzia percorsi per l'innovazione musicale, mentre i generi più isolati possono indicare nicchie stabili. Comprendere queste posizioni strategiche può informare le decisioni dell'industria musicale riguardo a opportunità di collaborazione e previsione di tendenze.

Nonostante queste intuizioni, l'analisi presenta limitazioni. I dati sono limitati a Spotify e al periodo 2000-2020, escludendo tendenze recenti e altre piattaforme.

L'assegnazione di un singolo genere per artista, sebbene utile per la semplificazione, potrebbe non cogliere la complessità degli artisti multi-genere, introducendo un bias. La tendenza di Spotify a promuovere generi popolari potrebbe anche aver distorto la percezione della centralità di generi meno diffusi.

In sintesi, questo progetto ha dimostrato il valore dell'analisi di rete per decifrare le complesse relazioni nell'industria musicale. Fornisce una solida base per future esplorazioni delle dinamiche che modellano le collaborazioni musicali contemporanee.