

# Analisi dell'affinità musicale degli artisti di Last.fm

Calogero Giudice  
c.giudice@studenti.unipi.it  
Student ID: 530155

## ABSTRACT

Questo progetto analizza una rete di artisti musicali del social network *last.fm* in base a dei coefficienti di similarità (Sample e Jaccard Ratios) che potrebbero rappresentare delle affinità a livello stilistico o di pubblico. Sono state esaminate sia le **caratteristiche strutturali** della rete, procedendo a svolgere una previsione, in base ai Ratios menzionati in precedenza e al coefficiente di popolarità, su possibili aumenti o cali di popolarità futuri.

Questi risultati possono rappresentare un **modello** che potrebbe riflettersi nello scenario musicale odierno, fornendo nuovi studi e strumenti che possono essere utili a prevedere eventuali nuovi scenari. Infatti, possono esserci artisti emergenti che potrebbero godere di maggiore popolarità futura se ben collegati, con artisti di popolarità alta e collegati con ratio alti; allo stesso tempo, si potrebbero notare alcuni artisti ritenuti popolari che potrebbero avere un trend opposto e quindi in declino.<sup>1</sup>

## KEYWORDS

Social Network Analysis, musica, *Last.fm*, affinità, similarità, popolarità

## ACM Reference Format:

Calogero Giudice. 2025. Analisi dell'affinità musicale degli artisti di Last.fm. In *Analisi dell'affinità musicale degli artisti di Last.fm*. ACM,

### <sup>1</sup>Project Repositories

Data Collection: [https://github.com/Kalo9603/2024\\_Giudice/tree/main/data\\_collection](https://github.com/Kalo9603/2024_Giudice/tree/main/data_collection)

Network Analysis: [https://github.com/Kalo9603/2024\\_Giudice/tree/main/network\\_analysis](https://github.com/Kalo9603/2024_Giudice/tree/main/network_analysis)

Open Problem: [https://github.com/Kalo9603/2024\\_Giudice/tree/main/open\\_problem](https://github.com/Kalo9603/2024_Giudice/tree/main/open_problem)

Report: [https://github.com/Kalo9603/2024\\_Giudice/tree/main/report](https://github.com/Kalo9603/2024_Giudice/tree/main/report)

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).  
*SNA '25, 2024/25, University of Pisa, Italy*

© 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM...\$0.00

<https://doi.org/10.1145/nnnnnnn.nnnnnnn>

New York, NY, USA, 3 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

## 1 INTRODUZIONE

Nell'epoca dell'affermazione dell'industria musicale digitale, piattaforme come *last.fm* offrono tanti dati sugli artisti musicali, sulle abitudini di ascolto degli utenti e sulle relazioni tra i singoli profili musicali (gli utenti, gli artisti, gli album, etc.). Gli artisti presenti nella piattaforma sono stati analizzati sotto una prospettiva di rete sociale: si possono ipotizzare così degli scenari, dei fenomeni interessanti: la similarità tra artisti, la popolarità degli stessi, l'affermazione di artisti emergenti o l'effetto contrario.

L'analisi parte dalla costruzione di una rete basata su coefficienti di similarità; di questi ne sono stati individuati due:

- Il *Sample Ratio*, ovvero l'intensità del legame tra due artisti in base ai simili condivisi.
- Lo *Jaccard Ratio*, ovvero l'affinità relativa all'insieme totale di simili.

Matematicamente, è molto più probabile che lo *Jaccard* sia inferiore del rispettivo *Sample*, poiché il denominatore del primo è sempre maggiore del secondo (a meno che tali denominatori non coincidono).

In parallelo si è stimata la **popolarità** degli artisti presenti nel dataset: per far ciò ci si è basati sullo **Z-Index**, che mette in relazione il numero di ascoltatori e quello di ascolti. Ciò ha permesso di classificare gli artisti in tre classi: **mainstream**, **medio** ed **emergente**; a questi ne sono stati affiancati altri tre: **possibile mainstream**, **possibile medio** e **ritorno emergente** in base al trend di (de)crescita calcolato.

L'obiettivo principale della seconda parte progetto è quello di identificare le caratteristiche strutturali degli artisti mainstream e di quelli emergenti ed, eventualmente, di prevedere un cambio di status in base a determinate metriche.

## 2 DATA COLLECTION

La raccolta dei dati da analizzare si è concentrata su tre fasi specifiche:

- (1) La generazione di un elenco di artisti in base agli artisti simili;
- (2) La creazione di una lista di archi che collegano due artisti se entrambi i coefficienti di similarità superano

la soglia del 50%. Il *Sample Ratio* è calcolato rispetto al numero di simili condivisi, mentre lo *Jaccard Ratio* rispetto all'insieme totale dei simili;

- (3) Il successivo accostamento del numero di ascolti, di ascoltatori e il calcolo delle popolarità degli artisti.

Gli elenchi degli artisti e degli artisti simili sono stati ottenuti per mezzo dell'API di *last.fm* e, tramite programmazione ad oggetti, sono state generate le altre liste. Nel complesso sono stati generati quattro file:

- *artists.csv*, una lista di 28 870 artisti;
- *links.csv*, un elenco di 98 890 archi che collegano 11 745 artisti (il 40.68% del totale). A causa delle lunghe tempistiche è stato generato il dataset sul 64.26% dei possibili archi totali (oltre 416 milioni);
- *popularity.csv*: la medesima lista di *artists.csv* con inclusi il numero di ascoltatori, quello di ascolti, il coefficiente di popolarità e il suo logaritmo;
- *popularity\_z.csv*: la medesima lista di *popularity.csv* a cui si accoda lo Z-Index.

In particolare, i coefficienti di correlazione citati sono calcolati come segue.

| Sigla               | Formula  |
|---------------------|--|
| Z-Index             | $Z_i = \frac{pop_{log,i} - \mu}{\sigma}$                         |
| Media               | $\mu = \frac{1}{N} \sum_{i=1}^N pop_{log,i}$                     |
| Deviazione standard | $\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (pop_{log,i} - \mu)^2}$ |
| Sample Ratio        | $SR(a, b) = \frac{ S_a \cap S_b }{k}$                            |
| Jaccard Ratio       | $JR(a, b) = \frac{ S_a \cap S_b }{ S_a \cup S_b }$               |

**Table 1: Formule utilizzate per la classificazione e l'analisi strutturale**

### Dati selezionati

Al fine di avere un dataset iniziale di artisti si è pensato, partendo da quattro artisti provenienti da quattro generi musicali diversi ( Annalisa per il pop italiano; Giorgio Vanni per le sigle, Katy Perry per la quota internazionale e Ado per il mercato giapponese).

Quindi, partendo da essi, viene generata la lista di artisti simili a cominciare da quello in input: se non è presente, l'artista trovato viene aggiunta in lista. La medesima operazione viene svolta, alla prima iterazione, sui primi quattro artisti; poi, dalla seconda, si considerano anche tutti quelli aggiunti in coda. In questo modo è stato ottenuto un dataset

di 28 870 artisti.

Al termine si è passati, basandosi sul dataset degli artisti, alla generazione del dataset dei *links* del grafo.

**Crawling Methodology and Assumptions.** Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum<sup>2</sup>.

- (1) item 1
- (2) item 2
- (3) item 3

### 3 NETWORK CHARACTERIZATION

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

#### Comparision with ER

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

#### Comparision with BA

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

$$\lim_{n \rightarrow \infty} x = \sum_{i \in B} \frac{1}{2} \quad (1)$$

<sup>2</sup>Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur [www.abcd.com](http://www.abcd.com).

#### 4 TASK: OPEN QUESTION

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

#### 5 DISCUSSION

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.