

Giulia Castagnacci, 581749
Elisabetta Giordano, 536265
Gruppo: casta_gio

Analisi Tendenze Musicali

Progetto Finale Big Data

Link repository GitHub: <https://github.com/GiuliaCastagnacci/BigData2.git>

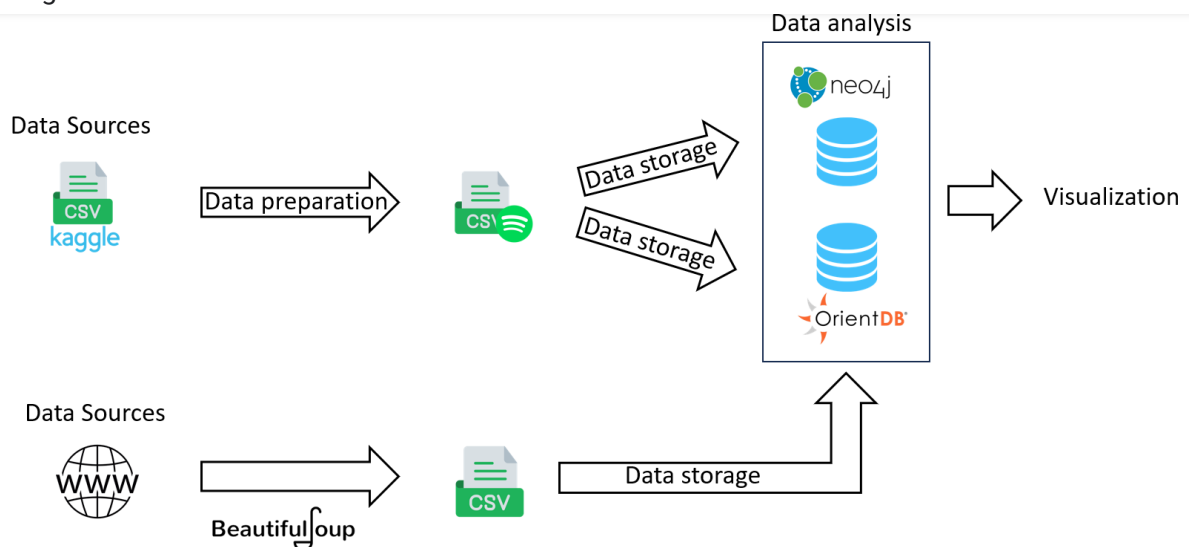
Obiettivo

Realizzazione di due Knowledge Graph :

- il primo per l'analisi di tendenze musicali; in particolare, analizzando come valenza, energia e acustica variano nel corso degli anni;
- il secondo per rappresentare gli avvenimenti storici in determinati anni, al fine di vedere se esiste una correlazione tra brano ed evento.

Viene inoltre effettuato un confronto tra le due tecnologie Neo4j e OrientDB nella gestione di “big data graphs” evidenziando le differenze tra i due sistemi.

Di seguito lo schema dell'architettura:



Dataset utilizzati

Dataset Spotify:

Per poter effettuare l'analisi delle tendenze musicali si è scelto di utilizzare il seguente dataset disponibile sul sito di Kaggle:

<https://www.kaggle.com/datasets/yamaerenay/spotify-dataset-19212020-600k-tracks?select=tracks.csv>

Il dataset iniziale tracks.csv contiene i dati relativi ai brani caricati su Spotify dal 1921 al 2020 ed è formato da 20 colonne e 586655 righe.

A partire da questo dataset è stata inizialmente effettuata un'operazione di filtraggio delle righe poichè la versione free di Neo4J ha un limite massimo di nodi e relazioni che può gestire. Per questo motivo si è scelto di eliminare alcuni record secondo i seguenti criteri:

1. per ogni anno sono stati eliminati i brani con popolarità uguale a 0, in quanto non significativi
2. effettuando un conteggio di quanti nodi (n) potessero essere supportati in neo4j, per ogni anno sono stati selezionati i brani con popolarità maggiore, cercando di mantenere sempre una certa omogeneità nel numero di brani per ciascun anno.

In particolare definito $n/\text{numero degli anni} = \text{numeroBraniMassimo}$:

- a. se un anno possiede un numero di brani maggiore di *numeroBraniMassimo* allora vengono selezionati solo i primi *numeroBraniMassimo* con popolarità più alta.
- b. se un anno possiede un numero di brani minore di *numeroBraniMassimo* allora vengono mantenuti tutti i brani relativi a quell'anno.

Successivamente sono state selezionate le colonne significative per la realizzazione del nostro Progetto:

- name: nome del brano
- release_date: anno di pubblicazione del brano
- energy: quanto è energica una canzone nell'intervallo da 0 a 1
- acousticness: quanto è acustica una traccia nell'intervallo da 0 a 1
- valence: la positività della traccia nell'intervallo da 0 a 1

Pertanto, il dataset finale ottenuto è formato da 5 colonne e 149353 righe.

	release_date	name	valence	energy	acousticness
0	1922	Carve	0.127	0.445	0.674
1	1922	Rum and coca cola	0.941	0.652	0.532
2	1922	Nuits De Chine	0.400	0.288	0.994
3	1922	Tu Verras Montmartre	0.897	0.280	0.995
4	1922	I'll see you in my dreams	0.481	0.391	0.996
5	1923	Heebie Jeebies	0.590	0.232	0.995
6	1923	Song of the Islands	0.381	0.130	0.976
7	1923	Hotter Than That	0.854	0.321	0.973
8	1923	You Can't Shush Katie	0.675	0.231	0.995
9	1923	Anybody Here Want to Try My Cabbage?	0.580	0.201	0.976

Dataset Eventi Storici:

Il dataset è stato creato estraendo i dati presenti nella pagina web riportata di seguito, contenente gli eventi globali più importanti che si verificano ogni anno dal 1920 al 2020, attraverso l'utilizzo di BeautifulSoup:

<https://eu.usatoday.com/story/money/2020/09/06/the-worlds-most-important-event-every-year-since-1920/113604790/>

Si è posta l'attenzione sui dati relativi agli anni e gli eventi corrispondenti.

Successivamente è stata effettuata una pulizia dei campi relativi alla colonna eventi.

	year	event
0	1920	Women's Suffrage
1	1921	Chinese Communists Rise
2	1922	British Empire Shrinks
3	1923	Great Kanto Earthquake
4	1924	From Lenin to Stalin
5	1925	Scopes Monkey Trial
6	1926	U.S. Starts Numbered Highway System
7	1927	Lindbergh Nonstop to Paris
8	1928	Earhart Crosses Atlantic
9	1929	Wall Street Crashes

Neo4J

Importazione dati e creazione grafo:

Per l'importazione dei dati, in Neo4j è molto semplice, basta richiamare il comando LOAD. Poiché il formato del dataset che abbiamo a disposizione è di tipo CSV, bisogna specificarlo.

```
LOAD CSV WITH HEADERS FROM 'file:///spotify_filtered_update.csv' AS row
```

Dopo aver caricato i dati, ci concentriamo nella creazione del grafo.

Creiamo un nodo "Tutti gli anni"

```
CREATE (:AllYears {name: "Tutti gli anni"})
```

Creiamo un nodo "year" per ogni anno presente nel dataset, eliminando i duplicati.

```
LOAD CSV WITH HEADERS FROM 'file:///spotify_filtered_update.csv' AS row  
WITH DISTINCT row.release_date AS releaseDate  
CREATE (year:Year { year: toInteger(releaseDate)})
```

Collegiamo il nodo "Tutti gli anni" con una relazione "CONTIENE" con gli "year" creati.

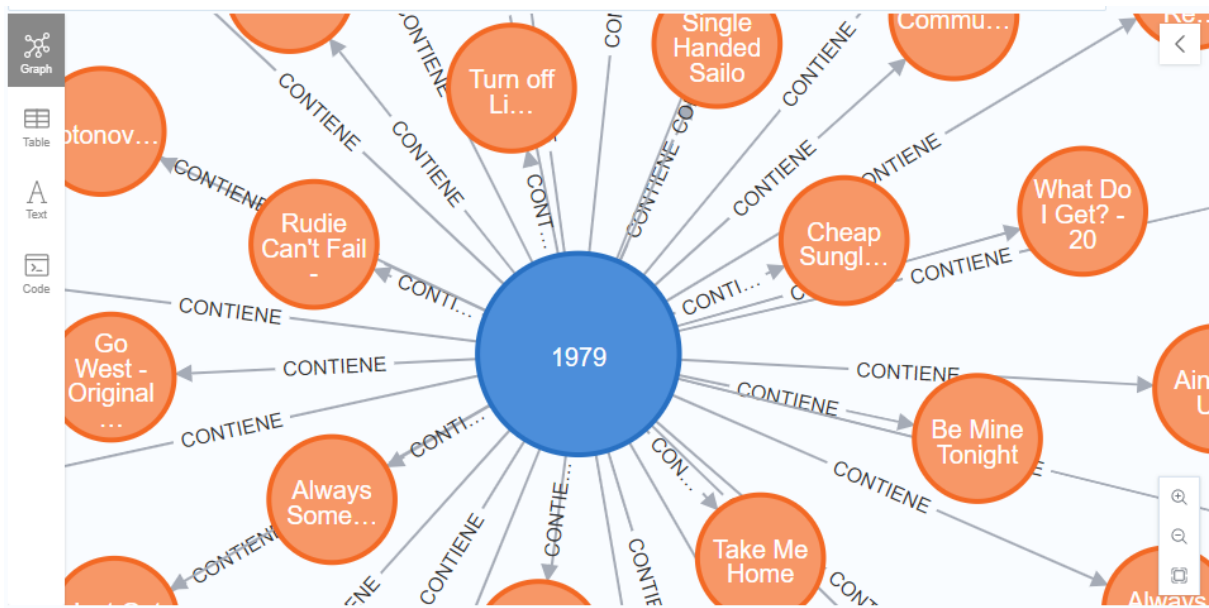
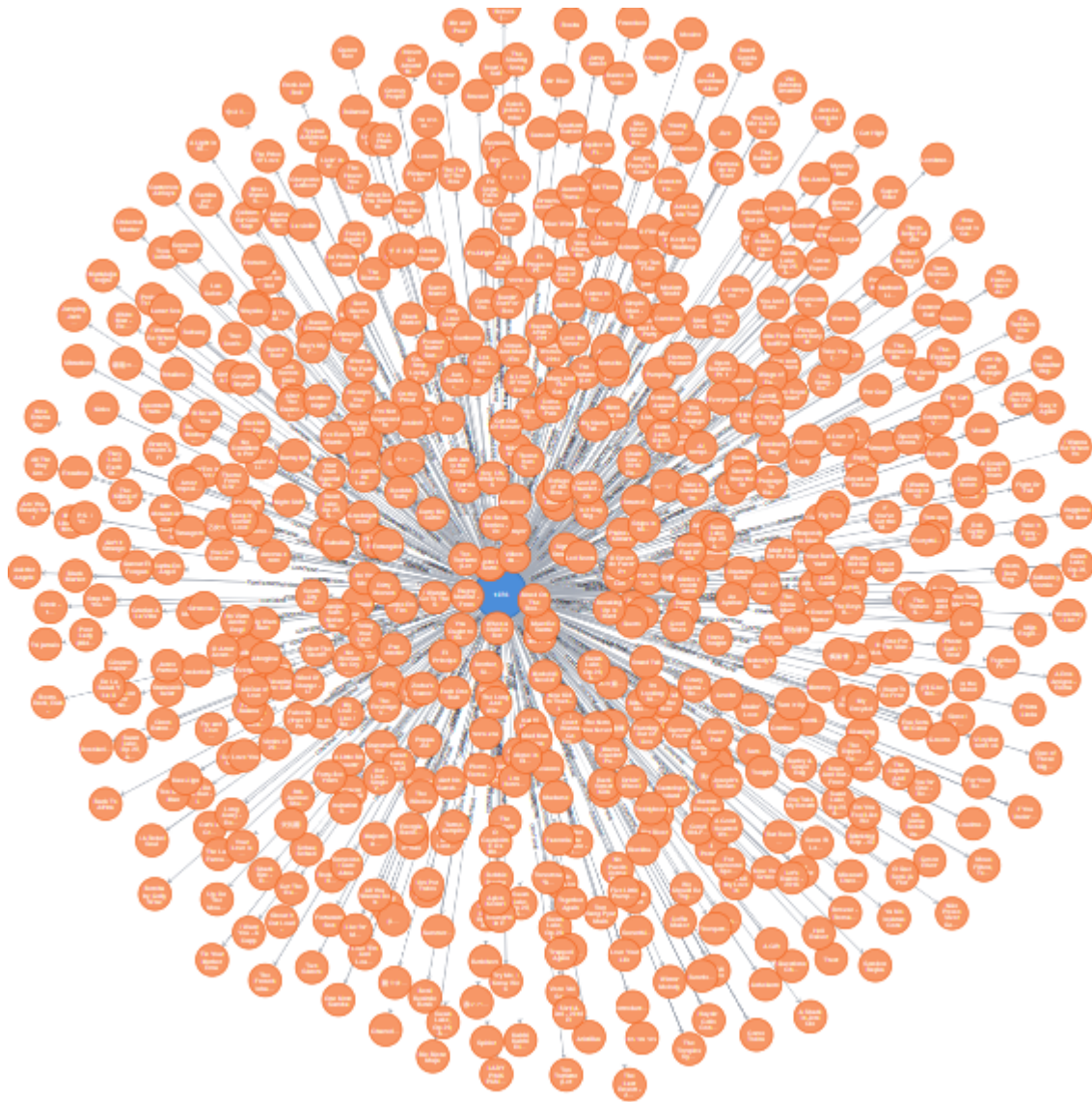
```
MATCH (allYears:AllYears {name: "Tutti gli anni"})  
MATCH (year:Year)  
MERGE (allYears)-[:CONTIENE]->(year)
```

Creiamo i nodi "track" che identificano i brani presenti. Li associamo all'anno corrispondente e settiamo gli attributi energy, valence e acousticness.

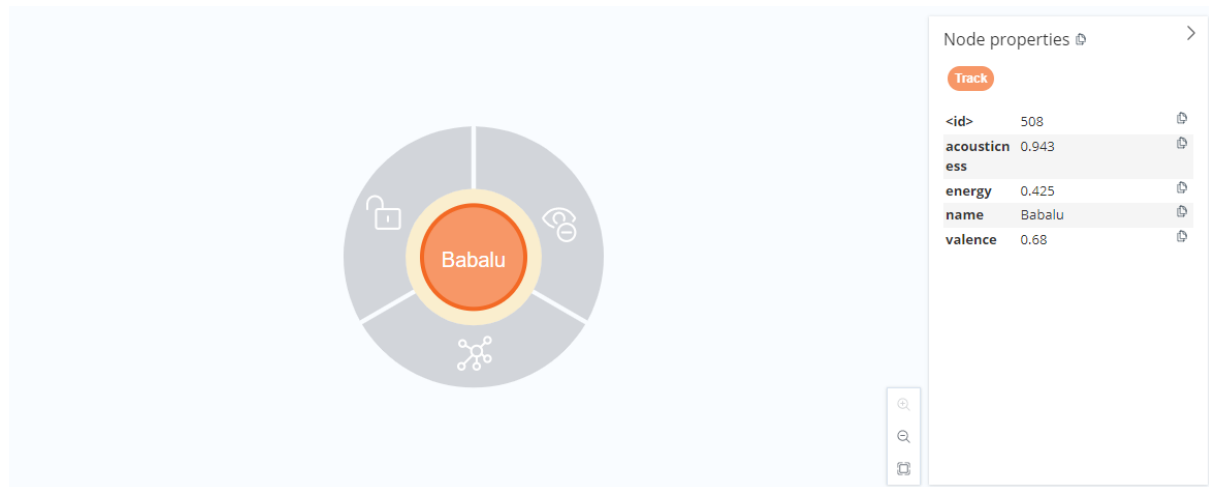
```
LOAD CSV WITH HEADERS FROM 'file:///spotify_filtered_update.csv' AS row  
MATCH (year:Year {year: toInteger(row.release_date)})  
CREATE (track:Track {name: row.name})  
MERGE (year)-[:CONTIENE]->(track)  
SET track.energy = toFloat(row.energy),  
    track.acousticness = toFloat(row.acousticness),  
    track.valence = toFloat(row.valence)
```

query time per creazione grafo : completato dopo 15823 ms.

Vediamo di seguito un esempio di visualizzazione di un anno con i relativi brani.



Dall'immagine si può notare come siano state assegnate valenza, acustica ed energia ad un brano.



Un procedimento analogo è stato seguito per l'importazione e la creazione del grafo degli eventi storici:

Importazione dati del dataset eventi storici e creazione nodi time, e event.

```
LOAD CSV WITH HEADERS FROM 'file:///datasetEventiStorici.csv' AS row
CREATE (:Time {time: row.time})
CREATE (:Event {event: row.event})
```

Creazione relazione ACCADE tra anno ed evento.

```
MATCH (t:Time {time: row.time})
MATCH (e:Event {event: row.event})
CREATE (t)-[:ACCADE]→(e)
```

Creazione nodo "all times" e relazione tra "all times" e "time"

```
CREATE (:AllTimes {name: 'all times'})

MATCH (at:AllTimes {name: 'all times'})
MATCH (t:Time)
CREATE (at)-[:INCLUDE]→(t)
```

query time per creazione grafo : completato dopo 406 ms.

Di seguito viene riportata la visualizzazione del grafo.

OrientDB

Importazione dati e creazione schema:

Per l'importazione di dati da un file CSV in un database OrientDB è necessario utilizzare un file di configurazione ETL. Questo file, in formato JSON, contiene informazioni su come e dove reperire i dati, come trasformarli e come caricarli nel database OrientDB.

```
{
  "config": {
    "log": "info",
    "credentials": {
      "user": "root",
      "password": "password"
    }
  },
  "source": {
    "file": {
      "path": "/OrientDB/datasets/spotifyDataset.csv"
    }
  },
  "extractor": {
    "csv": {}
  },
  "transformers": [
    {
      "vertex": {
        "class": "Branco",
        "fields": {
          "name": "${input.name}",
          "valence": "${input.valence}",
          "energy": "${input.energy}",
          "acoustiness": "${input.acoustiness}"
        }
      }
    },
    {
      "edge": {
        "class": "Rilasciato nel",
        "joinFieldName": "release_date",
        "lookup": "Anno.year",
        "direction": "out"
      }
    }
  ]
}
```



```

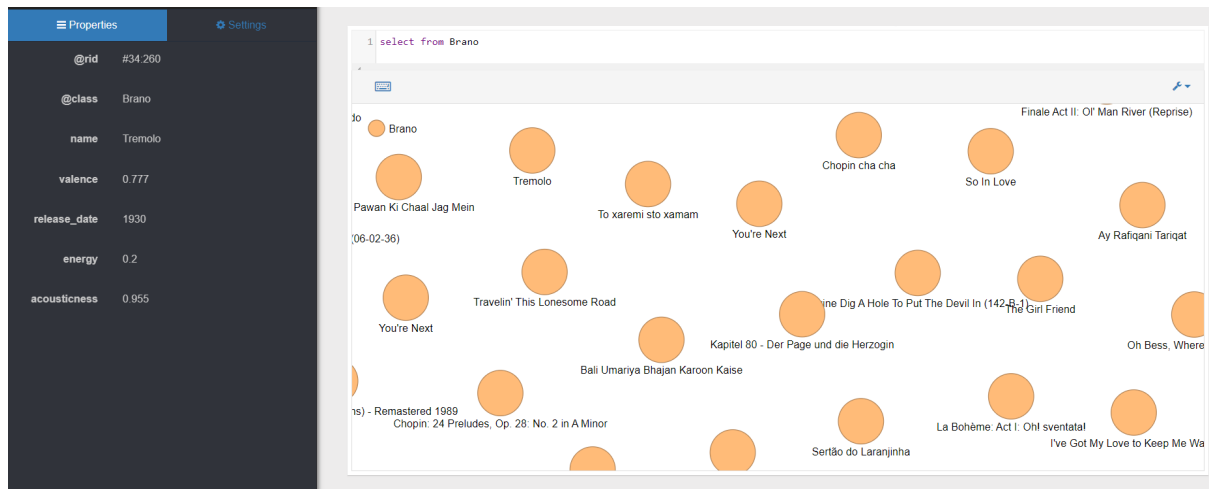
"loader": {
  "orientdb": {
    "dbURL": "plocal:/OrientDB/databases/ProgettoBD_spotify",
    "dbUser": "admin",
    "dbPassword": "admin",
    "dbType": "graph",
    "classes": [
      {"name": "Anno", "extends": "V"},
      {"name": "Brano", "extends": "V"},
      {"name": "Rilasciato nel", "extends": "E"}
    ],
    "indexes": [
      {
        "class": "Anno",
        "fields": ["year:integer"],
        "type": "UNIQUE"
      },
      {
        "class": "Brano",
        "fields": ["name:string", "valence:float", "energy:float", "acousticness:float"],
        "type": "NOTUNIQUE"
      }
    ]
  }
}

```

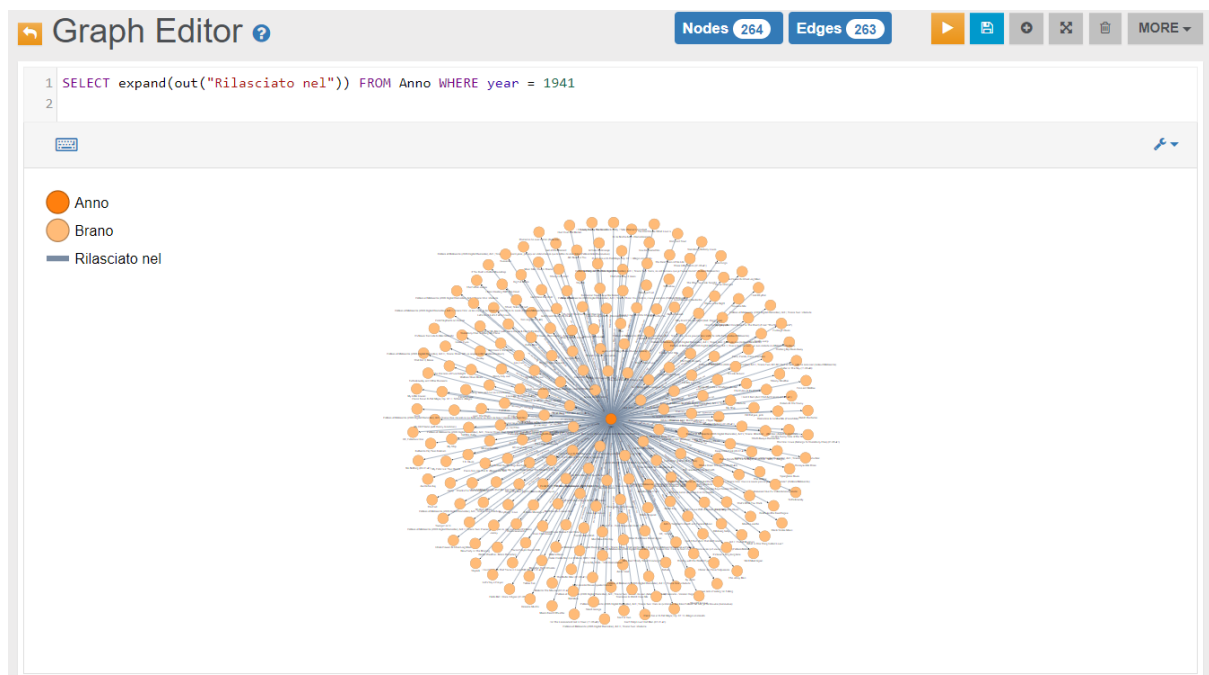
Le diverse sezioni del file di configurazione includono:

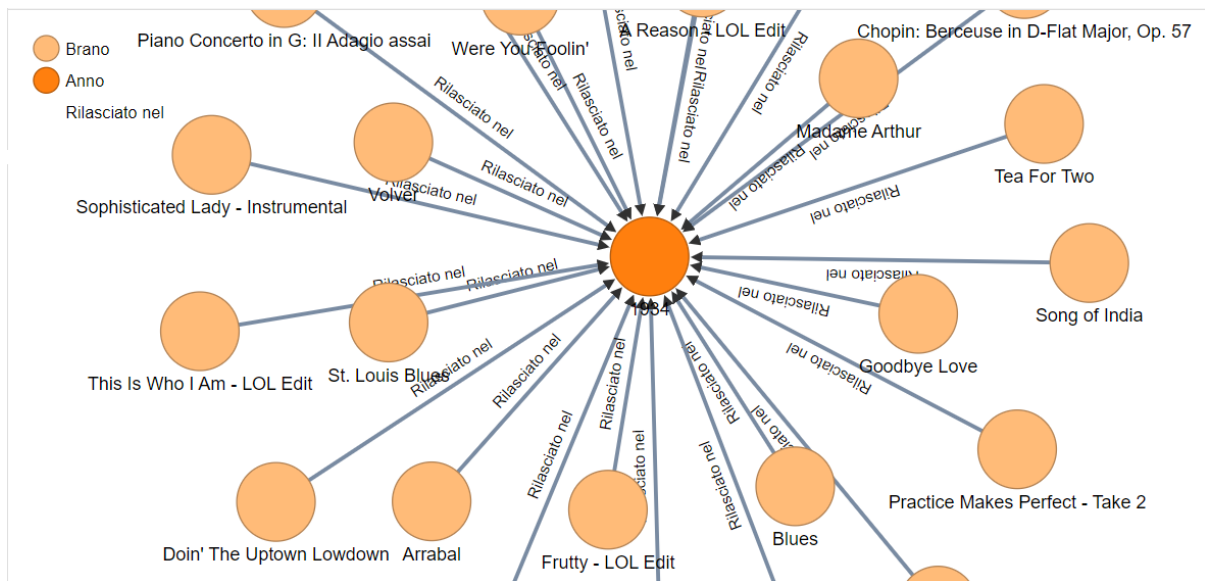
1. **"config"**: Questa sezione contiene le configurazioni globali per il processo ETL. In questo caso, ci sono due chiavi:
 - a. **"log"**: Specifica il livello di dettaglio dei log per il processo
 - b. **"credentials"**: Contiene le credenziali per accedere al database OrientDB.
2. **"source"**: Questa sezione specifica la sorgente dei dati che verranno importati. In questo caso, i dati provengono da un file CSV situato in una directory di OrientDB.
3. **"extractor"**: Questa sezione specifica come estrarre i dati dalla sorgente. Viene utilizzato un estrattore CSV.
4. **"transformers"**: Questa sezione definisce le trasformazioni da applicare ai dati prima del caricamento nel database OrientDB. Ci sono due trasformazioni:
 - a. La prima trasformazione crea nodi di tipo "Brano" nel database OrientDB con campi "name", "valence", "energy", e "acousticness" ottenuti dai dati di input.
 - b. La seconda trasformazione crea archi di tipo "Rilasciato nel" che collegano i nodi di tipo "Anno" ai nodi di tipo "Brano". Questo viene fatto utilizzando la chiave "release_date" nei dati di input per cercare l'anno corrispondente nel database OrientDB e collegarlo al brano.
5. **"loader"**: Questa sezione definisce come caricare i dati trasformati nel database OrientDB. Le chiavi principali sono:
 - a. **"dbURL"**: Specifica l'URL del database OrientDB a cui si sta accedendo.
 - b. **"dbType"**: Specifica il tipo di database, che è "graph" in questo caso.
 - c. **"classes"**: Definisce le classi che verranno utilizzate nel database OrientDB, inclusi "Anno", "Brano" e "Rilasciato nel".
 - d. **"indexes"**: Specifica gli indici da creare nel database OrientDB. In questo caso, vengono definiti indici univoci per la classe "Anno" basati sul campo "year" e indici non univoci per la classe "Brano" basati su vari campi come "name", "valence", "energy" e "acousticness".

Visualizzazione:



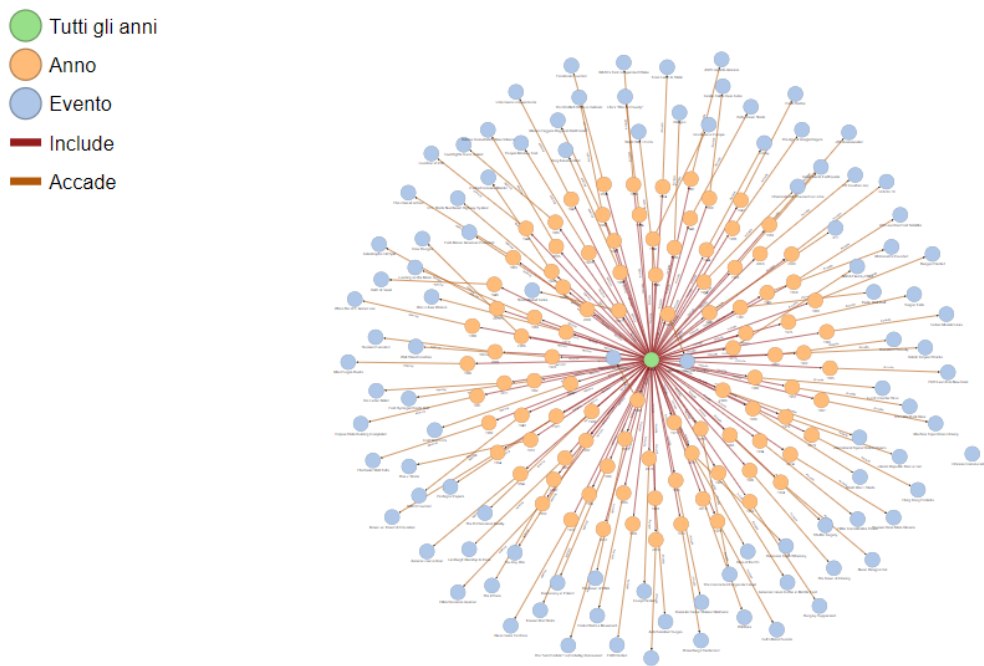
Dall'immagine possiamo notare come le proprietà nome, valenza, energia e acustica, vengano caricate e assegnate correttamente al nodo Brano.

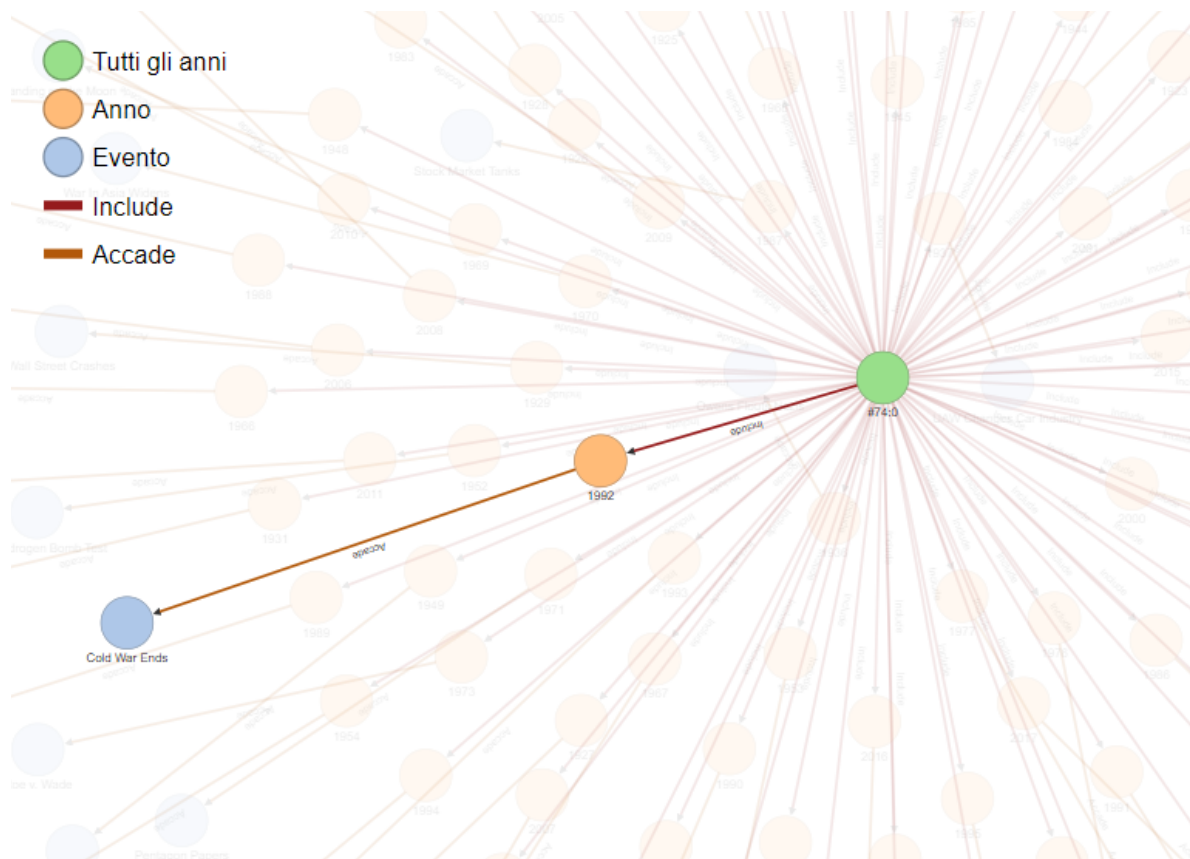




Dall'immagine possiamo notare come venga assegnata correttamente la relazione “Rilasciato nel” tra il nodo Brano e il nodo “Anno”, corrispondente all’anno di pubblicazione del brano.

Lo stesso procedimento viene effettuato per il dataset relativo agli eventi storici. Di seguito un esempio di visualizzazione:





In questo caso, è stato creato un nodo “Tutti gli anni” che include gli anni in cui si verificano gli eventi storici e un nodo “Anno” che è collegato, attraverso la relazione “Accade”, all’evento corrispondente.

In conclusione, i nodi e le relazioni create sono:

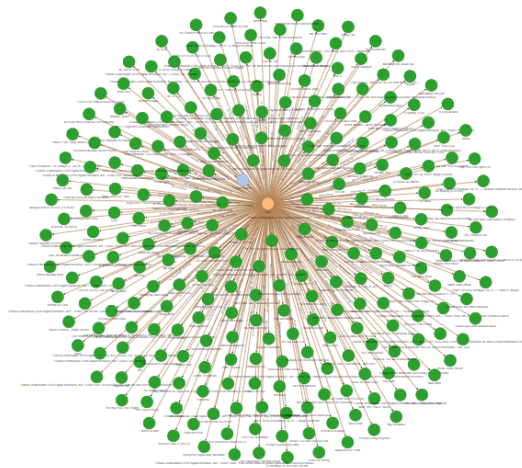
Vertex Classes

Name ?	Color	SuperClasses ?	Alias	Abstract	Clusters ?	Default Cluster	Cluster Selection ?	Records
Anno		V		<input type="checkbox"/>	[34, 35, 36, 37, 38, 39, 40, 41]	34	round-robin ▾	100
Brano		V		<input type="checkbox"/>	[26, 27, 28, 29, 30, 31, 32, 33]	26	round-robin ▾	149,352
Evento		V		<input type="checkbox"/>	[42, 43, 44, 45, 46, 47, 48, 49]	42	round-robin ▾	101
Tutti gli anni		V		<input type="checkbox"/>	[74, 75, 76, 77, 78, 79, 80, 81]	74	round-robin ▾	1
V				<input type="checkbox"/>	[10, 11, 12, 13, 14, 15, 16, 17]	10	round-robin ▾	149,554

Edge Classes

Name ?	Color	SuperClasses ?	Alias	Abstract	Clusters ?	Default Cluster	Cluster Selection ?	Records
Accade		E		<input type="checkbox"/>	[58, 59, 60, 61, 62, 63, 64, 65]	58	round-robin ▾	99
E				<input type="checkbox"/>	[18, 19, 20, 21, 22, 23, 24, 25]	18	round-robin ▾	149,551
Include		E		<input type="checkbox"/>	[82, 83, 84, 85, 86, 87, 88, 89]	82	round-robin ▾	100
Rilasciato nel		E		<input type="checkbox"/>	[50, 51, 52, 53, 54, 55, 56, 57]	50	round-robin ▾	149,352

● Brano
● Evento
● Anno
— Accade
Rilasciato nel



Analisi effettuate

Caso di studio 1:

Inizialmente si è posta l'attenzione sulle tre variabili di interesse, valence, energia e acustica, su come sono gli andamenti e se c'è o meno una correlazione tra loro.

Vediamo come:

- Valence: rappresenta la positività o la negatività dell'atmosfera di una canzone.
 - Valori più alti di valence indicano una maggiore positività e gioia.
 - Valori più bassi di valence possono indicare toni più tristi o malinconici.
- Energy: rappresenta il livello di attività e intensità di una canzone.
 - Valori più alti di energy indicano canzoni più dinamiche ed energetiche.
 - Valori più bassi di energy possono rappresentare canzoni più tranquille o rilassanti.
- Acousticness: rappresenta la presenza di strumenti acustici nella canzone rispetto agli strumenti elettronici.
 - Valori più alti di acousticness indicano una maggiore presenza di elementi acustici, come chitarre o strumenti a corda.
 - Valori più bassi di acousticness possono indicare una maggiore presenza di strumenti elettronici o di produzione digitale.

La media dei valori di "valence", "energy" e "acousticness" può quindi fornire indizi sullo stato emotivo e sullo stile delle canzoni.

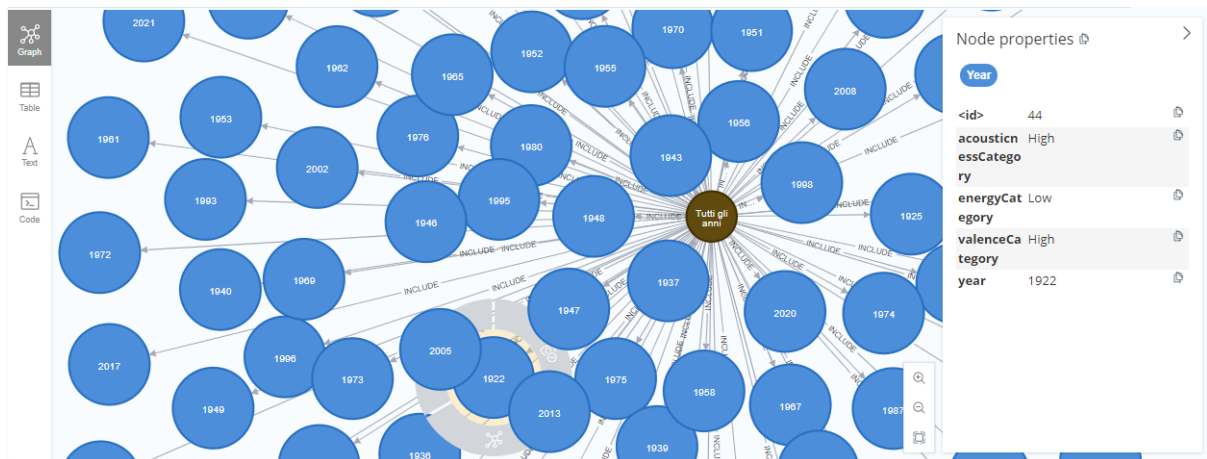
Alcuni dei valori ottenuti sono:

PROPERTIES			
year	avg_valence	avg_energy	avg_acousticness
1922	0.52	0.33999999999999997	0.76
1923	0.5601941747572814	0.17475728155339793	0.8932038834951466
1924	0.61	0.2342857142857142	0.87999999999999992
1925	0.6317689530685928	0.2433212996389893	0.8758122743682347
1926	0.4170157068062834	0.1548429319371716	0.5900523560209394
1927	0.5471544715447155	0.30162601626016283	0.864227642276424
1928	0.3683274021352309	0.16370106761565859	0.8889679715302531
1929	0.5575342465753427	0.22191780821917792	0.8479452054794512
1930	0.5236956521739127	0.24065217391304342	0.8615217391304315
1931	0.5448051948051951	0.24155844155844153	0.8402597402597419

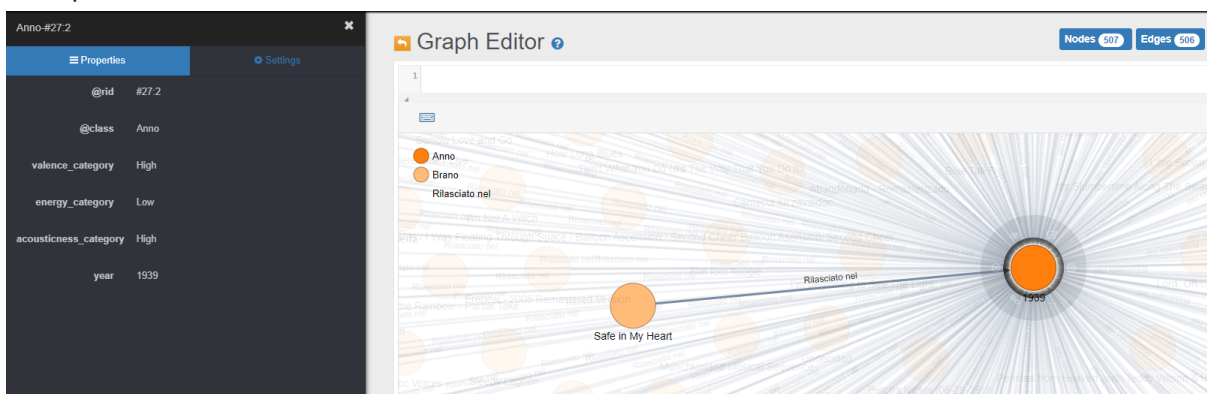
A partire dal risultato ottenuto sono state definite per ciascun anno tre nuove proprietà: avg_valence, avg_energy, avg_acousticness e due valori “High” e “Low”. In particolare, verrà assegnato il valore “High” alla proprietà se la media ottenuta è > 0.5, “Low” altrimenti.

Vediamo come vengono rappresentate graficamente le proprietà aggiunte nei due GDBMS:

Esempio di visualizzazione in Neo4J:

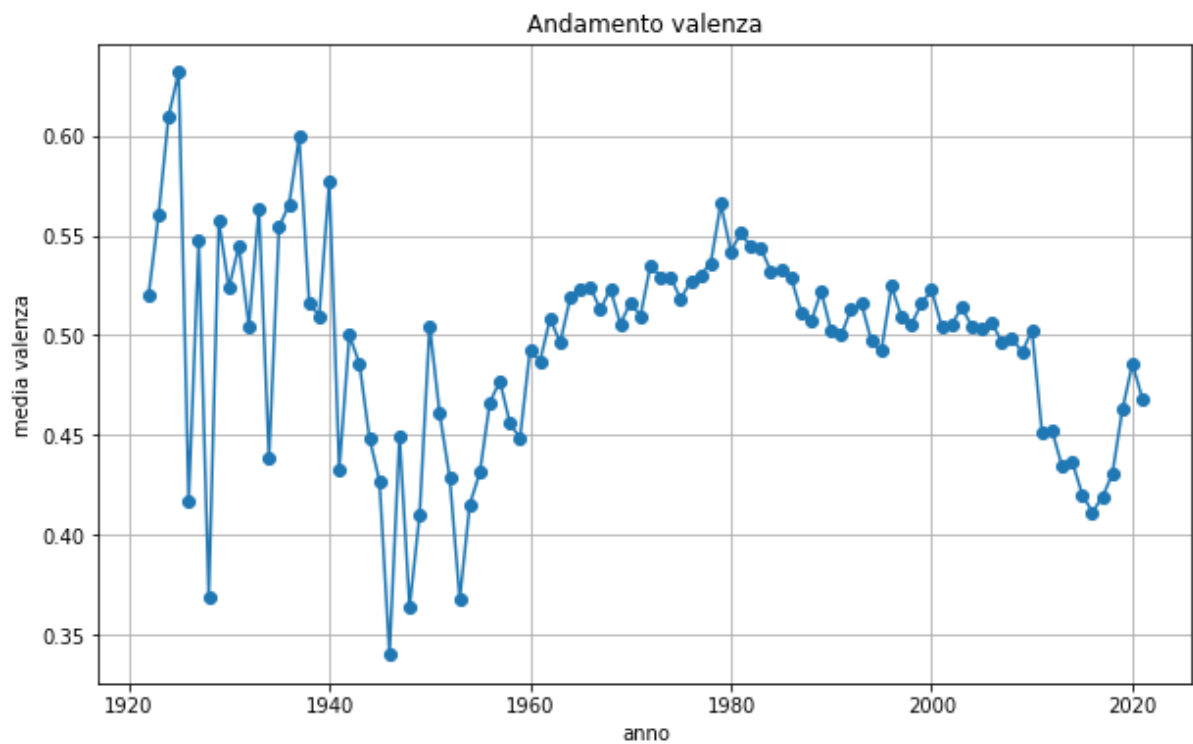


Esempio di visualizzazione in OrientDB:

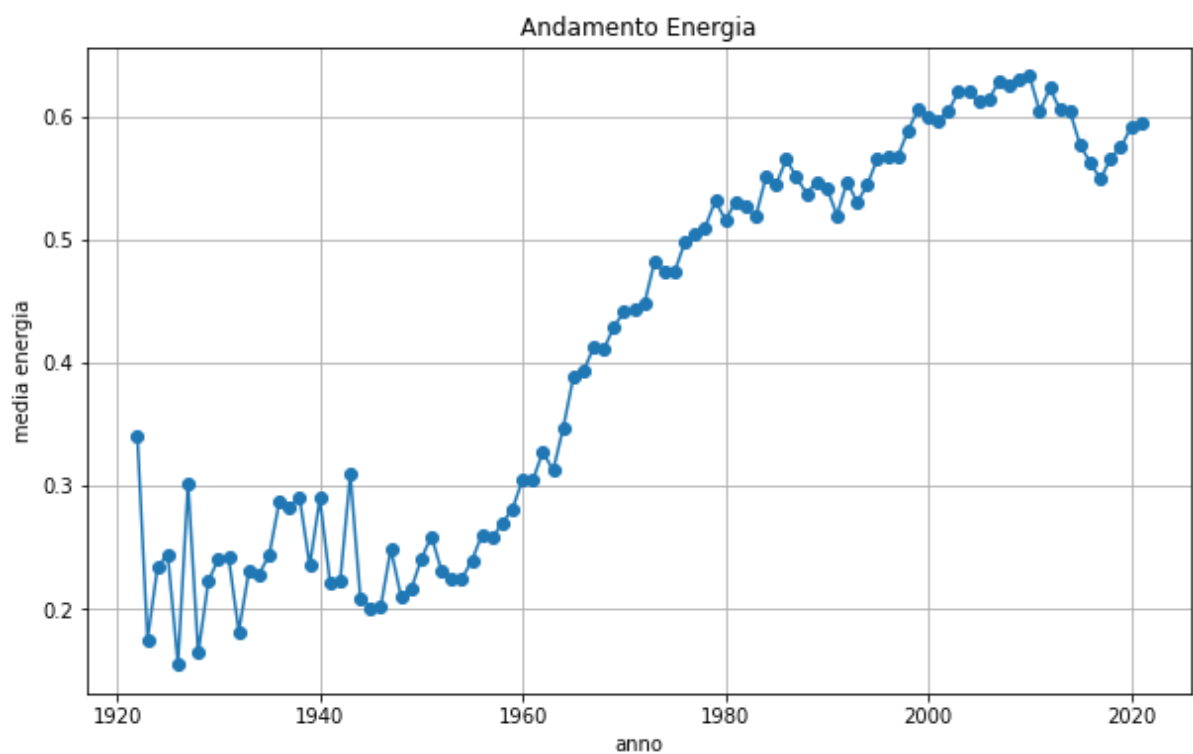


A seguito dei risultati ottenuti si può visualizzare l'andamento delle tre variabili negli anni

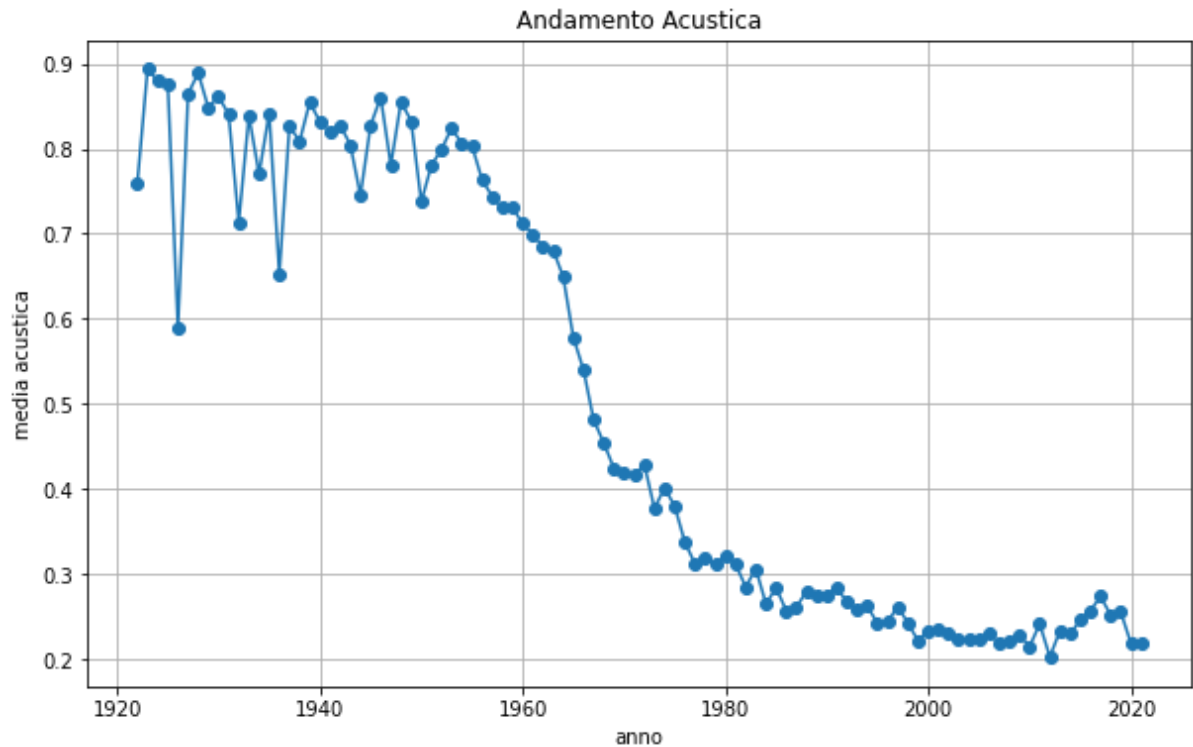
Tendenza Valenza:



Tendenza Energia:



Tendenza Acustica:



Come si può notare dai tre grafici Energia e Acustica hanno un andamento speculare, mentre la Valenza non presenta similarità significative rispetto all'andamento delle altre due variabili, in alcuni anni si presenta uguale all'Energia, in altri uguale all'Acustica, ma non è presente una proporzionalità definita.

Non essendoci una similarità netta tra Energia e Valenza, si può notare come un brano 'positivo' non debba necessariamente avere una maggiore energia, come invece si potrebbe pensare.

Caso di studio 2:

Query sui brani che possiedono caratteristiche specifiche richieste dall'utente. In questo modo l'utente sarà in grado di visualizzare brani che corrispondono alle sue preferenze musicali.

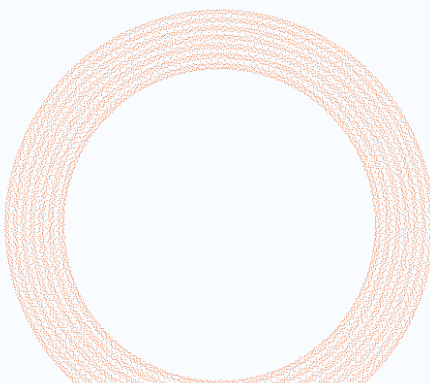
Query 1: Visualizzazione dei brani ritenuti positivi

Visualizzazione in Neo4J:

```

1 MATCH (track:Track)
2 WHERE track.valence > 0.5
3   AND track.acousticness > 0.5
4   AND track.energy > 0.5
5 RETURN track;
6
7

```



Overview

Node labels

* (11649) Track (11649)

Displaying 11,649 nodes, 0 relationships.

```

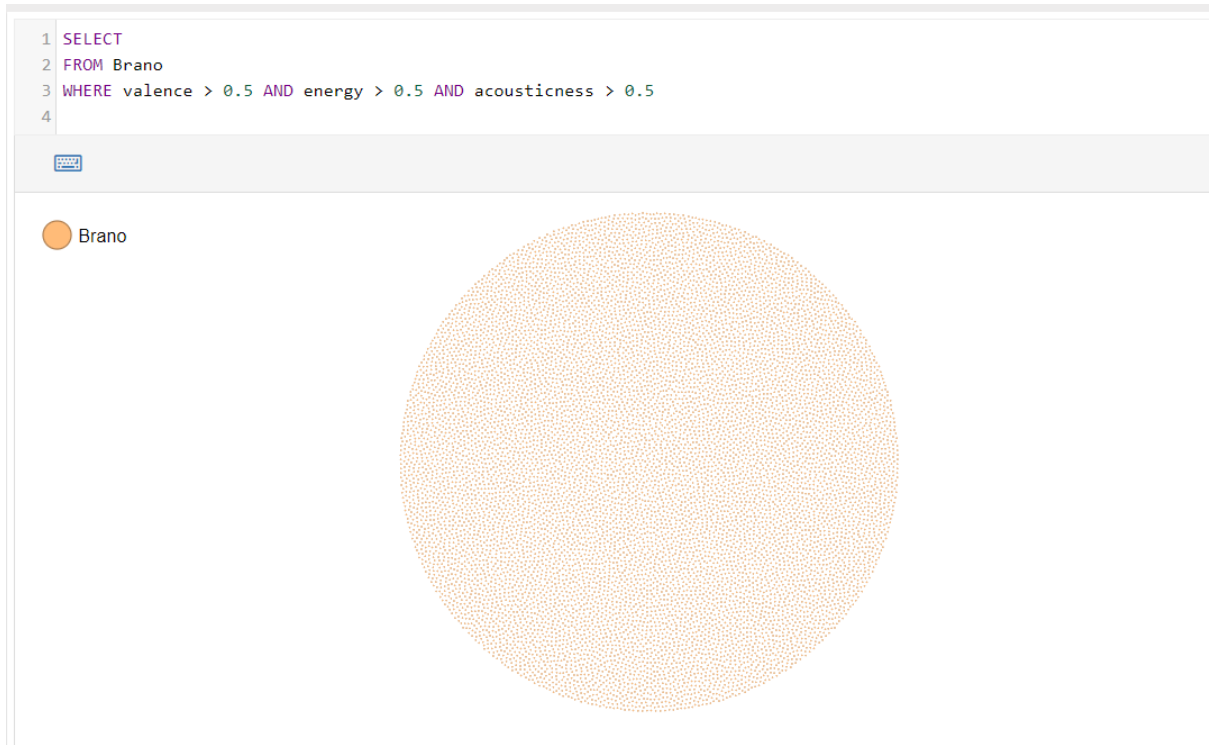
1 MATCH (track:Track)
2 WHERE track.valence > 0.5
3   AND track.acousticness > 0.5
4   AND track.energy > 0.5
5 RETURN track;
6
7

```

track
(:Track {acousticness: 0.784,name: "Ritmo Tropical",valence: 0.903,energy: 0.784})
(:Track {acousticness: 0.786,name: "Cuca",valence: 0.9,energy: 0.664})
(:Track {acousticness: 0.889,name: "Yellow Fire",valence: 0.789,energy: 0.597})
(:Track {acousticness: 0.779,name: "Aaj Ki Raat Kuchh Hogi Aisi Baat",valence: 0.556,energy: 0.576})
(:Track {acousticness: 0.583,name: "Junto al Río Magdalena (Rumba Gitana)",valence: 0.792,energy: 0.527})
(:Track {acousticness: 0.774,name: "Suavecito",valence: 0.927,energy: 0.669})

Query completata dopo 2.84 s.

Visualizzazione in OrientDB:



COMMAND

SELECT name AS brano FROM Brano WHERE valence > 0.5 AND energy > 0.5 AND acousticness > 0.5

PROPERTIES

brano
Lonely Melody
My Pretty Girl
Ory's Creole Trombone
Flagler Drive
Doctor Jazz
Moonrise On the Lowlands
Black Bottom Stomp - 1992 Remastered
She's Long, She's Tall
Daddy and Home
Mulskinners Blues

1 2 3 4 5 6 7

10 25 50 100 1000 5000

Query executed in 0.58 sec. Returned 11649 record(s). Limit: 200000 (CHANGE IT)

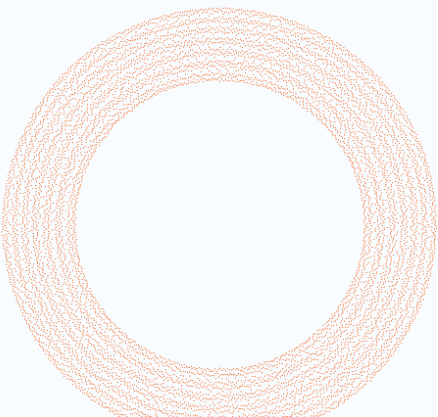
Table Raw Explain

Query executed in 0.58 sec. Returned 11649 record(s)

Query 2: Visualizzazione dei brani ritenuti negativi

Visualizzazione in Neo4J:

```
1 MATCH (track:Track)
2 WHERE track.valence < 0.5
3   AND track.acousticness < 0.5
4   AND track.energy < 0.5
5 RETURN track;
```



Overview

Node labels

- (8197)
- Track (8197)

Displaying 8,197 nodes, 0 relationships.

```
1 MATCH (track:Track)
2 WHERE track.valence < 0.5
3   AND track.acousticness < 0.5
4   AND track.energy < 0.5
5 RETURN track;
```

track
(:Track {acousticness: 0.486,name: "Get Out Of Town - Remastered - April 1992",valence: 0.4,energy: 0.411})
(:Track {acousticness: 0.409,name: "The Rules of the Road",valence: 0.373,energy: 0.268})
(:Track {acousticness: 0.347,name: "I Got Rhythm",valence: 0.405,energy: 0.228})
(:Track {acousticness: 0.242,name: "Nuestra Decisión",valence: 0.392,energy: 0.476})
(:Track {acousticness: 0.453,name: "To Be Forgiven",valence: 0.364,energy: 0.451})
(:Track {acousticness: 0.298,name: "What's Another Year",valence: 0.185,energy: 0.28})

Query completata dopo 2.80 s.

Visualizzazione in OrientDB:



COMMAND

SELECT name AS brano FROM Brano WHERE valence < 0.5 AND energy < 0.5 AND acousticness < 0.5

PROPERTIES

brano
Words (Between the Lines of Age) - 2009 Remaster
Impressioni di Settembre
Five Years - 2012 Remaster
Old Man - 2009 Remaster
Melissa
Changes
Roma capoccia
Como Vai Você - Versão Remasterizada
The Ballad Of El Goodo
San Vicente

21 22 23 24 25 26 27

10 25 50 100 1000 5000

Query executed in 0.514 sec. Returned 8197 record(s). Limit: 200000 [\(CHANGE IT\)](#)

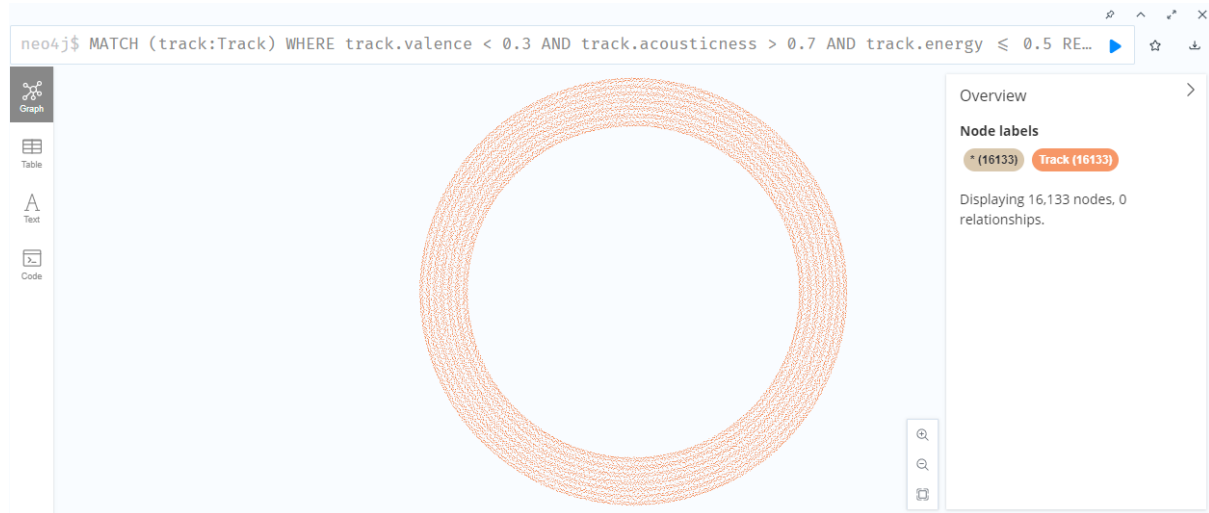
Table Raw Explain

Query executed in 0.514 sec. Returned 8197 record(s).

Come si può vedere dai risultati delle query, all'interno del dataset, sono presenti più brani con caratteristiche 'positive' piuttosto che 'negative'.

Query 3: Visualizzazione dei brani consigliati all'utente: valenza < 0.3, energia ≤ 0.5, acustica > 0.7

Visualizzazione in Neo4J:

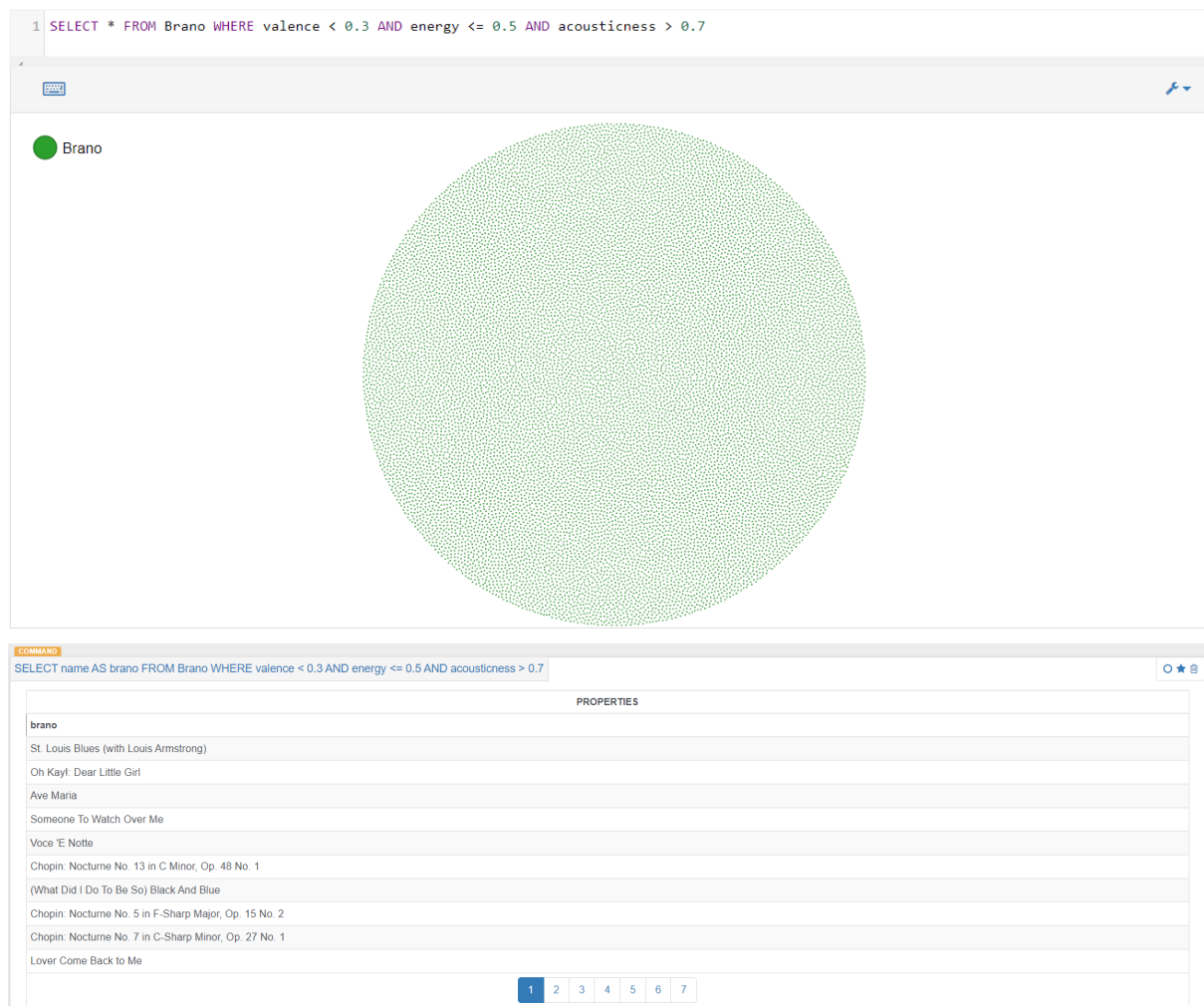


```
1
2 MATCH (track:Track)
3 WHERE track.valence < 0.3
4   AND track.acousticness > 0.7
5   AND track.energy ≤ 0.5
6 RETURN track;
```

track
(:Track {acousticness: 0.995,name: "Squeeze Me",valence: 0.219,energy: 0.121})
(:Track {acousticness: 0.889,name: "Felléas et Mélisande (2006 Digital Remaster), Act II, Scene Two: Ah! Ah! Tout va bien, cela ne sera rien (Golaud/Mélisande)",valence: 0.216,energy: 0.0673})
(:Track {acousticness: 0.973,name: "Felléas et Mélisande (2006 Digital Remaster), Act II, Scene One: C'est au bord d'une fontaine aussi qu'l vous a trouvée? (Felléas/Mélisande)",valence: 0.207,energy: 0.0793})
(:Track {acousticness: 0.969,name: "Felléas et Mélisande (2006 Digital Remaster), Act II, Scene Two: Tiens, où est l'anneau que je t'avais donné? (Golaud/Mélisande)",valence: 0.296,energy: 0.15})
(:Track {acousticness: 0.958,name: "I Got It Bad (And That Ain't Good)",valence: 0.238,energy: 0.072})

Query completed after 4.35 s.

Visualizzazione in OrientDB:



Query executed in 0.582 sec. Returned 16133 record(s).

Inoltre, sono state fatte ulteriori due query :

1. valenza = 1, energia = 1, acustica = 1, per considerare un brano con un massimo valore di positività per tutte le caratteristiche;
2. valenza = 0, energia = 0, acustica = 0, per considerare un brano con un massimo valore di negatività per tutte le caratteristiche

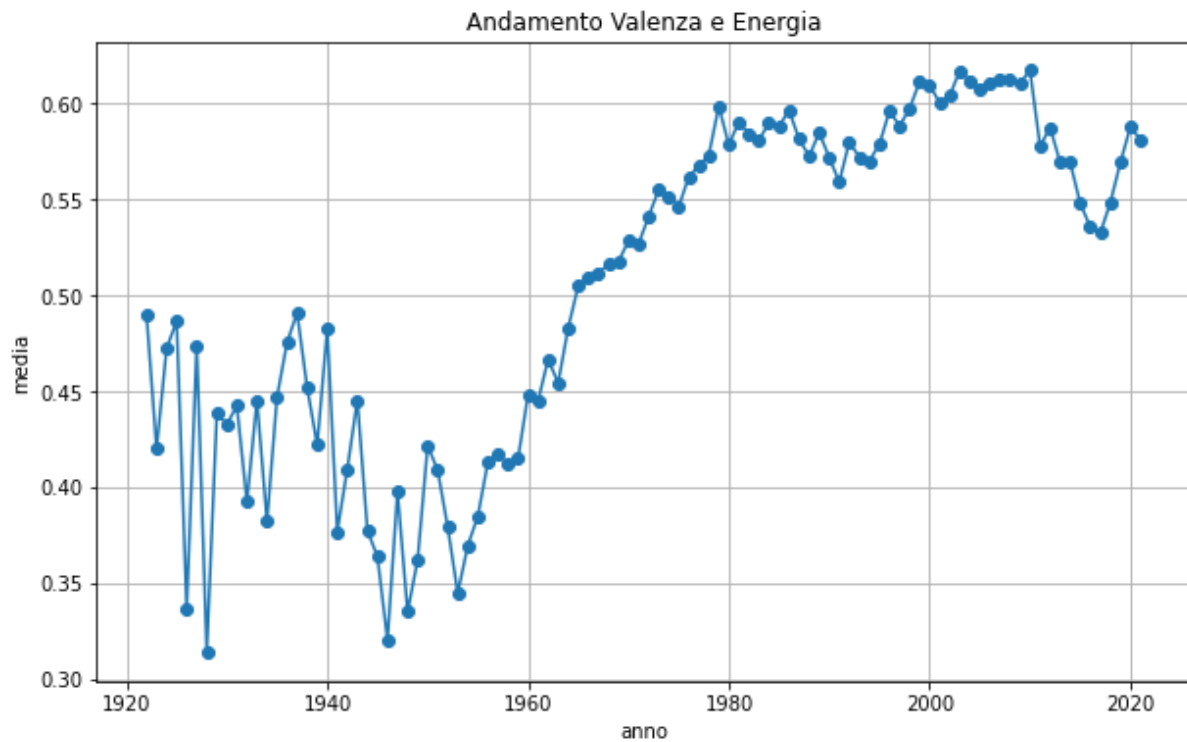
ed entrambe le query hanno prodotto un risultato nullo. Quindi non esiste un brano che rispecchia le caratteristiche minime o massime.

Caso di studio 3:

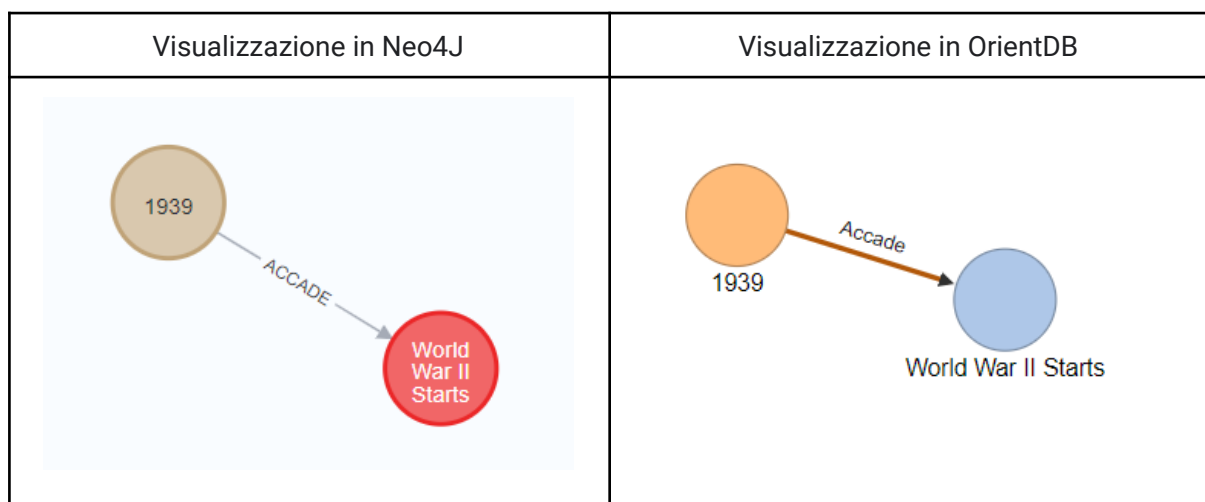
Dopo aver concluso le analisi delle tendenze musicali nel primo dataset, abbiamo deciso di ampliare ulteriormente l'analisi aggiungendo il secondo dataset *Eventi Storici* per vedere se le caratteristiche musicali fossero state influenzate anche dal periodo storico in cui venivano realizzate.

Si è scelto di porre l'attenzione su valenza ed energia perchè sono le due caratteristiche che rispecchiano al meglio lo stato d'animo del brano.

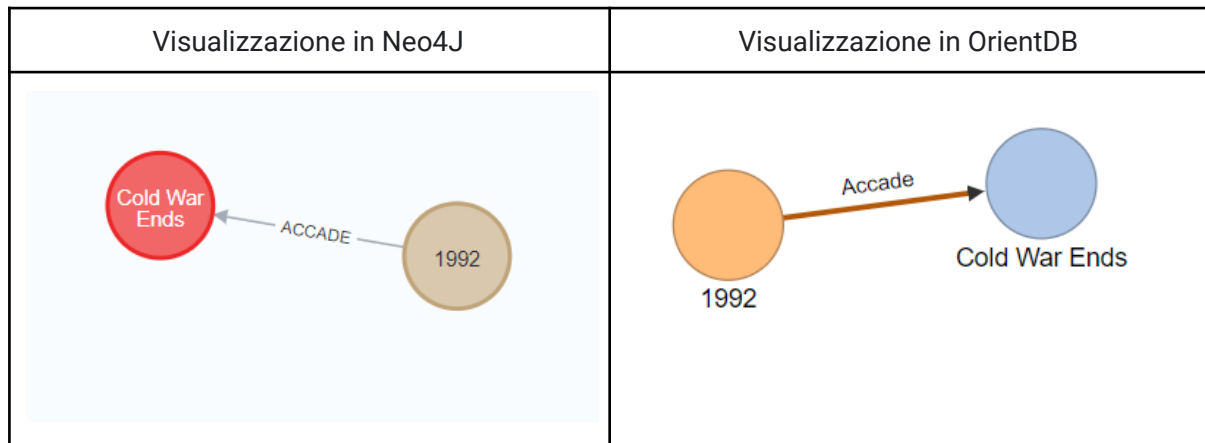
Di seguito viene riportato l'andamento della media di energia e valenza nel corso degli anni, considerando come 'positivo' un valore maggiore di 0.5, 'negativo' altrimenti:



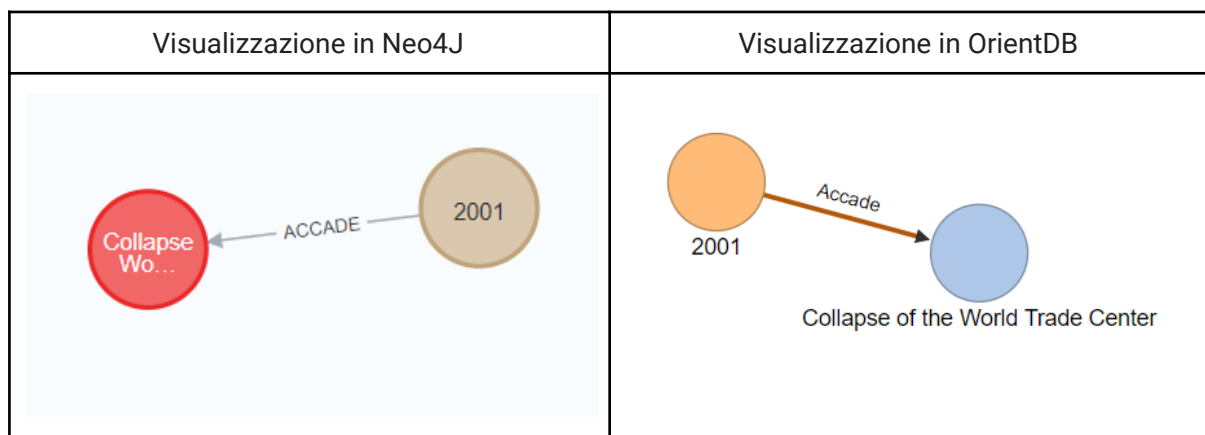
Query 1: evento 'negativo' → Inizio della Seconda Guerra Mondiale



Query 2: evento 'positivo' → Fine della Guerra Fredda



Query 3: nessuna corrispondenza → Attentato dell'11 settembre 2001



year	avg_valence_energy
1939	0.37223320158102774
1992	0.5301538461538458
2001	0.5504871794871788

(OrientDB) Query executed in 1.049 sec.

Come possiamo notare non in tutti i casi l'evento storico influisce le caratteristiche musicali dei brani.

Confronto tecnologie

Nella realizzazione dei casi d'uso sono state identificate le seguenti differenze nell'uso dei due GDBMS Neo4J e OrientDB:

Intuitività nella costruzione e l'importazione dei dati nel grafo:

- Neo4j: Neo4j è molto intuitivo nella costruzione del grafo grazie al suo modello di dati basato su nodi e relazioni. Creare e visualizzare grafi è relativamente semplice, specialmente con l'aiuto di strumenti di visualizzazione come Neo4j Browser. Lo stesso vale anche per l'importazione, che viene effettuata tramite il comando LOAD CSV FROM "path del file".
- OrientDB: OrientDB risulta meno intuitivo nell'importazione dei dati all'interno del grafo, poiché richiede l'uso di uno script esterno per caricare i dataset. Per quanto riguarda la creazione e la visualizzazione dei grafi è relativamente semplice grazie allo strumento di visualizzazione OrientDB Studio.

Linguaggio Query:

- Neo4j: Neo4j utilizza il linguaggio di query Cypher, che è stato progettato specificamente per interrogare grafi. Cypher offre una sintassi espressiva e intuitiva per eseguire query complesse su dati di grafo.
- OrientDB: OrientDB supporta SQL esteso per query di grafi. Questo significa che è possibile utilizzare SQL per interrogare il database, il che potrebbe essere più familiare per chi ha esperienza con SQL nei database relazionali.

Documentazione disponibile:

- Neo4j: Neo4j ha una vasta e ben documentata comunità di utenti e una vasta raccolta di risorse, tra cui documentazione ufficiale, guide, forum e tutorial. La documentazione di Neo4j è ampiamente disponibile e aggiornata.
- OrientDB: La documentazione di OrientDB è meno estesa rispetto a quella di Neo4j, ma comunque adeguata.

Tempi di prestazione:

- Nelle query eseguite, OrientDB ha tempi di esecuzione minori rispetto a Neo4j.

Conclusioni

Utilizzare un sistema di gestione di database grafico (GDBMS) per analizzare tendenze musicali e identificare correlazioni tra la musicalità di una canzone e gli eventi storici può essere estremamente utile per diverse ragioni:

- **Struttura flessibile:** Un GDBMS è progettato per rappresentare e gestire dati complessi con relazioni complesse, il che lo rende adatto per memorizzare informazioni sulla musicalità delle canzoni e sugli eventi storici in un formato facilmente accessibile.
- **Ricerca efficiente:** I GDBMS consentono di eseguire ricerche complesse e query grafiche per identificare correlazioni tra i dati. Questo significa si possono trovare facilmente le canzoni associate a eventi storici specifici o cercare modelli nella musicalità in base a periodi storici.
- **Analisi delle tendenze:** Utilizzando le capacità di analisi dei dati dei GDBMS, è possibile identificare tendenze musicali nel corso del tempo e vedere come queste tendenze possono essere correlate agli eventi storici, come periodi di guerra, cambiamenti culturali o innovazioni tecnologiche.
- **Visualizzazione dei dati:** I GDBMS consentono di creare visualizzazioni grafiche dei dati, che possono aiutare a comunicare le correlazioni e le tendenze in modo chiaro ed efficace.