

**UNIVERSIDAD PERUANA DE CIENCIAS APLICADAS**

ESCUELA DE POSTGRADO

PROGRAMA DE MAESTRÍA EN DATA SCIENCE

**TRABAJO FINAL**

Redes Complejas

**INTEGRANTES**

Valdéz Jara, Bryan Luis (E202210910)

Llanos Gutierrez, Jean Pierre (E201536465)

**PROFESOR**

Rebeca Lourdes Flores Atunga

**Lima, 26 de febrero del 2024**

## Índice

Contenido

[Índice 2](#_Toc159857729)

[Contenido 2](#_Toc159857730)

[**1.** **Descripción del problema o situación real.** 3](#_Toc159857731)

[**2.** **Adquisición y descripción del conjunto de datos (Dataset)** 4](#_Toc159857732)

[**3.** **Representación gráfica de los datos y métricas.** 5](#_Toc159857733)

[**4.** **Experimentos e interpretación de resultados** 6](#_Toc159857734)

[**4.1.** **Construcción de la red** 6](#_Toc159857735)

[**4.1.1.** **Conexión** 6](#_Toc159857736)

[**4.1.2.** **Recolección** 6](#_Toc159857737)

[**4.1.3.** **Creación de la red** 7](#_Toc159857738)

[**4.1.4.** **Análisis de centralidades** 8](#_Toc159857739)

[**4.1.5.** **Análisis PageRank** 9](#_Toc159857740)

[**4.1.6.** **Análisis de comunidades** 11](#_Toc159857741)

[**5.** **Conclusiones** 12](#_Toc159857742)

[**6.** **Referencias bibliográficas** 12](#_Toc159857743)

# **Descripción del problema o situación real.**

Para desarrollar el trabajo se utilizaron los datos de las canciones de las 50 Listas de reproducción más destacadas en Spotify, estos datos incluyen información de las canciones, características de audio, información del artista y su región como también el género al cual pertenece.

Estos datos nos ayudaran a explorar y comprender la estructura y las relaciones de las canciones en una red, basándonos en sus características para identificar canciones que están ganando popularidad e influencia en la red.

A continuación, algunos puntos clave sobre la importancia de la red en el contexto nacional:

* **Análisis de tendencia musical**

Identificando las canciones y géneros más influyentes en la región en términos de centralidad, se podría analizar tendencias en la música popular que se pueden aplicar a nivel nacional.

* **Segmentación de audiencias**

Identificar comunidades en las canciones populares que puede ser útil para dirigir campañas de marketing y recomendaciones a grupos de usuarios a nivel nacional

* **Descubrimiento de nuevas canciones**

Al analizar la red se puede identificar grupos de canciones similares o relacionadas a las preferidas por los usuarios, pero que aun no las hayan descubierto.

* **Recomendaciones de música**

Utilizando medidas de centralidad se puede identificar canciones clave que se relacionan o que actúan como puentes entre diferentes géneros, lo que nos permitiría tener recomendaciones más precisas.

# **Adquisición y descripción del conjunto de datos (Dataset)**

La red se generó utilizando servicios públicos de los servidores de Spotify, la información técnica se encuentra en el sitio <https://developer.spotify.com/> donde se obtuvieron las 50 listas de reproducción mas populares y con esta información se pudo acceder a los detalles de 3712 canciones, tomando en cuenta que estas listas de reproducción no tiene un límite establecido y muchas canciones se repiten en varias lista se realizó una limpieza de duplicados y nombres con caracteres diferentes, resultando un total de 2781 canciones filtradas a continuación, el detalle de la red Elegida.

N, K = G.order(), G.size()

avg\_deg = K / N

print("Number of nodes: ", N)

print("Number of edges: ", K)

print("Average degree: ", avg\_deg)

print("Directed: ",G.is\_directed())

* **Número de nodos: 2781**

Esto resultado significa que nuestra red tiene 2781 nodos, donde cada nodo representa una canción en Spotify

* **Número de conexiones: 268716**

Este dato nos indica la cantidad de conexiones 268716 entre las canciones de nuestra red

* **Grado promedio: 99.26**

Se refiere al número promedio de conexiones que tiene cada nodo en la red, para nuestros nodos apunta un promedio de 99, que significa que cada canción esta conectada a alrededor de 99 otras canciones.

* **Directo: False**

Significa que la red es no dirigida, las conexiones entre los nodos no tienen una dirección especifica asociada, indicando la relación simétrica entre canciones, es decir si una canción esta relacionada con otra, entonces la otra también esta relacionada con la primera.

# **Representación gráfica de los datos y métricas.**

Por el tamaño de los nodos, realzamos un subgrafo de 500 nodos para visualizar las relaciones

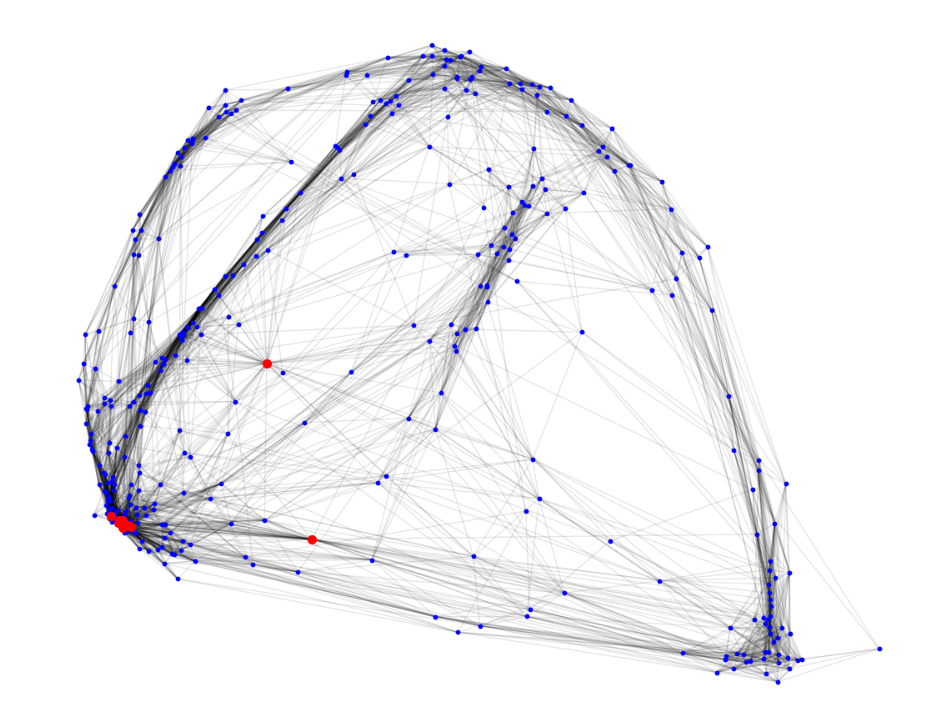
# Extrae un subgrafo con 1000 nodos aleatorios

subgrafo = nx.subgraph(G, list(G.nodes)[:500])

# Visualiza el subgrafo

nx.draw(subgrafo, with\_labels=False, node\_size=10, width=0.5)

plt.show()



**FIGURA 1 - SUBGRAFO DE LAS 10 CANCIONES MAS POPULARES**

# **Experimentos e interpretación de resultados**

# **Construcción de la red**

# **Conexión**

Para la construcción de la red se generaron un id y clave en la página principal de Spotify para conectarnos a los servicios de públicos de sus dominios a continuación el código donde se genera esta conexión.

#Conexión

client\_id = 'f0ee7d9fa4274eae8b08551b3140d843'

client\_secret = '5e3d306dd54e45e7b6aa94b3b55cd2e4'

auth\_manager = SpotifyClientCredentials(client\_id=client\_id, client\_secret=client\_secret)

sp = spotipy.Spotify(auth\_manager=auth\_manager)

# **Recolección**

Se realizo la consulta de 50 listas mas importantes en la región, este resultado dio un total de **3712** canciones con un identificador que se utilizó para obtener las características musicales, este proceso se realizo en bloques de 100 limitados por los mismos servicios de Spotify, al final generamos un archivo CSV con los atributos que consideramos relevantes para crear nuestra red.

#Lista de Playlist

playlist\_global = sp.featured\_playlists(limit=50)

len(playlist\_global['playlists']['items'])

#Obtener las canciones por playlist

list\_tracks = []

for item in playlist\_global['playlists']['items']:

  tracks = sp.playlist\_tracks(item['id'])

  for item in tracks['items']:

      track = item['track']

      list\_tracks.append({

          'track\_id': track['id'],

          'track\_name': track['name'],

          'artist\_id': [artist['id'] for artist in track['artists']],  # Lista de IDs de artistas

          'track\_popularity': track['popularity']

      })

# **Creación de la red**

Se utilizo las siguientes características musicales descritas a continuación:

* Bailable:

Es una medida de qué tan adecuada es una pista para bailar, basada en una combinación de elementos musicales como el ritmo, la estabilidad del tempo, la fuerza del ritmo y la regularidad general. Cuanto mayor sea el valor bailable, más adecuada será la pista para bailar

* Energía:

Representa la intensidad y actividad perceptiva de una pista. Las pistas de alta energía suelen ser rápidas, ruidosas y sonarán fuertes. La energía generalmente se refiere a la percepción de la actividad y la intensidad de la música, en lugar de ser una medida directa de los decibeles.

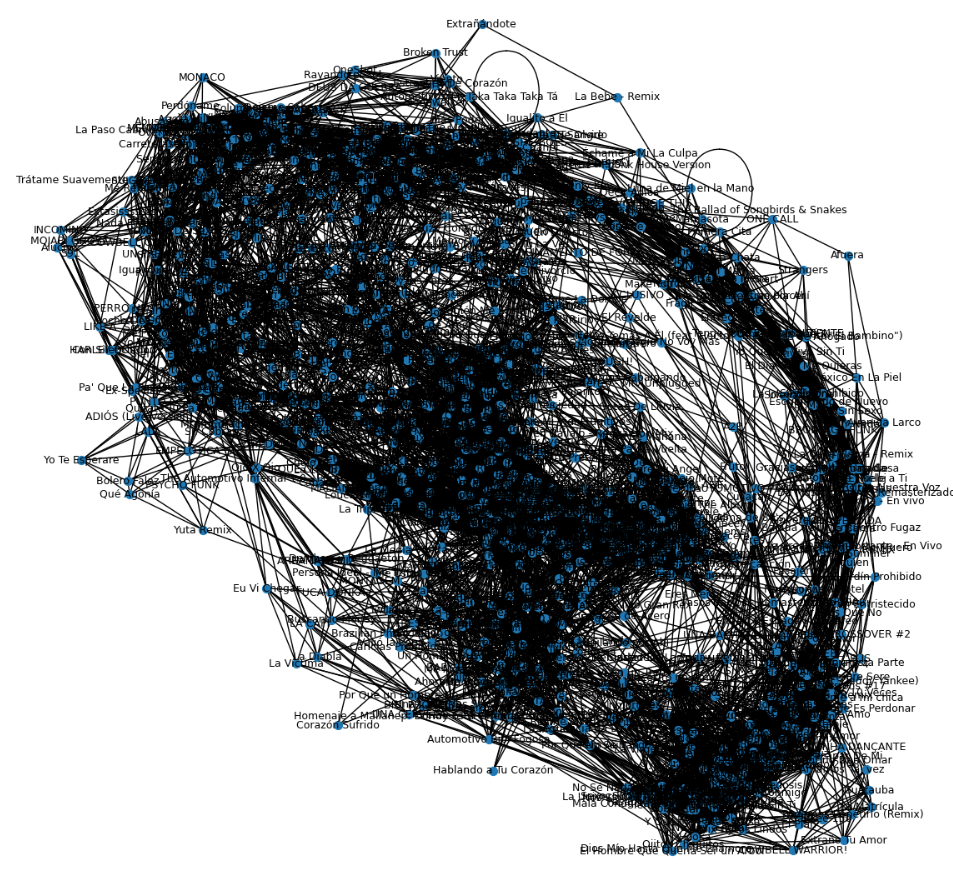
* Valencia:

Es una medida que describe la positividad musical transmitida por una pista. Las pistas con alta valencia suelen sonar más positivas (felices, alegres, eufóricas), mientras que las pistas con baja valencia tienden a sonar más negativas (tristes, melancólicas, enfadadas).

* Tempo:

Es simplemente la velocidad de la música, medida en beats por minuto (BPM). Un tempo alto significa que la música es más rápida, mientras que un tempo bajo indica que la música es más lenta.

Con estas características se procedió a normalizar y calcular la similitud de coseno, para armar nuestra red se utilizo como nodos las canciones y las relaciones de acuerdo a la similitud calculada



**FIGURA 2 - GRAFO CON LOS NODOS Y RELACIONES GENERADAS**

# **Análisis de centralidades**

Se realizo el cálculo y la identificación de los nodos centrales utilizando la métrica “Degree Centrality”

Gráfico, Gráfico de dispersión

Descripción generada automáticamente

**FIGURA 3 – Centralidad de grado**

Se identifico las canciones que tiene un alto grado de centralidad, visualizando la importancia dentro de la red de estas canciones, por su influencia estas canciones pueden ser utilizadas como referencias para las recomendaciones de música por su atractivo.



**FIGURA 4 – Lista de las 5 canciones con mayor centralidad por grado**

# **Análisis PageRank**

Se utilizo el algoritmo PageRank para identificar las canciones más importantes en función a su importancia en la red y como están relacionadas con otras canciones que también son importantes en la red.

# Calcular el PageRank de la red

pagerank = nx.pagerank(G)

# Imprimir el PageRank de cada nodo

for node, value in pagerank.items():

    print("PageRank de", node, ":", value)



**FIGURA 7 – Lista de canciones mas importantes en la red PageRank**

Imagen que contiene mapa, exterior, nieve, luz

Descripción generada automáticamente

**FIGURA 8 – Grafo canciones más importantes en la red PageRank**

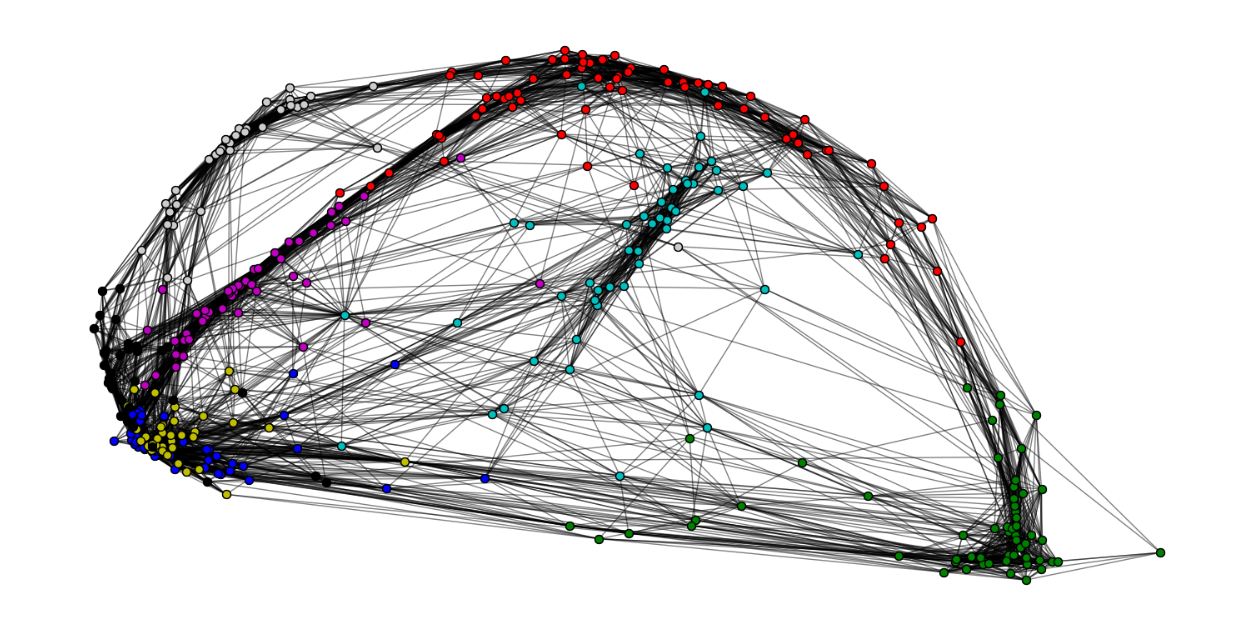
Papalote volando en el cielo

Descripción generada automáticamente con confianza media

**FIGURA 9 – Grafo canción ‘Necesito un amor’ top 1 PageRank**

# **Análisis de comunidades**

Utilizando el algoritmo de “louvain”, se pudo obtener un total de 8 comunidades



**FIGURA 5 – Grafo según comunidades**

Gráfico, Gráfico radial

Descripción generada automáticamente

**FIGURA 6 – Grafo de comunidades**

# **Conclusiones**

Mediante el presente Trabajo se pudo identificar las diez canciones más influyentes utilizando el análisis de PageRank de un total 2781 canciones

# **Referencias bibliográficas**