Project Unit 3

Subject: Social Networks Analysis

Student:
Luis Fernando Monterrubio Cota

Teacher:
Didier Gamboa

August 4, 2024

Diffusion of Information

I. Introduction

A. Overview

Urbano music, a genre that encompasses various styles such as reggaeton, Latin trap, and other urban Latin music genres, has surged in popularity over the last decade. Originating primarily in Latin American countries, urbano music blends rhythms and influences from hip-hop, dancehall, reggae, and Latin music. Its rise is marked by the success of artists like Bad Bunny, J Balvin, and Daddy Yankee, who have become global superstars. The genre's infectious beats, catchy melodies, and often provocative lyrics have resonated with a wide audience, propelling it to the forefront of the global music scene.

B. Purpose

The primary purpose of this network analysis is to map and visualize the collaborative relationships among top artists within the urbano genre. By constructing a network where nodes represent artists and edges represent collaborative tracks, we aim to uncover patterns and key players in the genre's ecosystem. Understanding these connections can provide insights into the collaborative nature of urbano music, identify central figures who play a significant role in the genre's evolution, and highlight how these artists influence each other's music.

C. Expected Insights

Through this analysis, we expect to uncover several insights, including:

- **The most central and influential artists within the urbano genre.**
- **Patterns of collaboration that may suggest sub-genres or stylistic similarities.**
- **The role of key collaborations in shaping the genre's trajectory.**

II. Mapping Process

A. Data Collection

The initial step in mapping the network involved gathering data on urbano artists and their collaborative tracks. This was accomplished using the Spotify Web API, a powerful tool for accessing Spotify's music data.

1) Authentication: Utilizing the spotipy library, we authenticated our access to the Spotify API using client credentials.
2) Artist Retrieval: Using the Spotify API, we queried and retrieved a list of top urbano artists. The goal was to gather a comprehensive list of influential artists within the genre.

```
def get_genre_artists(genre='urbano', total_artists=100):
    genre_artists = []
    limit = 50
    for offset in range(0, total_artists, limit):
        results = sp.search(q=f'genre:{genre}', type='artist', limit=limit, offset=offset)
        artists = [{'id': artist['id'], 'name': artist['name']} for artist in results['artists']['items']]
        genre_artists.extend(artists)
    return genre_artists

urbano_artists = get_genre_artists(total_artists=100)

```

3) Track Retrieval: For each artist, we retrieved a list of their tracks, focusing on those with collaborative features. This step involved handling pagination to ensure we collected a sufficient number of tracks for each artist.

```
def get_tracks(artist_name, total_tracks=100):
    tracks = []
    limit = 50
    for offset in range(0, total_tracks, limit):
        results = sp.search(q=f'artist:{artist_name}', type='track', limit=limit, offset=offset)
        tracks.extend(results['tracks']['items'])
        if not results['tracks']['next']:
```

```
                break
    return tracks

```

B. Data Processing

Once the raw data was collected, it required processing to extract meaningful information about collaborations:

1) Collaborations Extraction: For each track, we identified all featured artists. If a track featured multiple artists, we recorded these collaborations.

```
def get_collaborations_from_tracks(
    tracks, artist_name,
    top_artist_ids):
    data = []
    for track in tracks:
        track_id = track['id']
        track_name = track['name']
        for collaborator in track['
    artists']:
                collaborator_id =
    collaborator['id']
                collaborator_name =
    collaborator['name']
                if collaborator_name !=
     artist_name and collaborator_id
     in top_artist_ids:
                    data.append({
                        'artist_name':
    artist_name,
                        '
    collaborator_name':
    collaborator_name,
                        'track_name':
    track_name
                    })
    return data
```

2) Data Compilation: We compiled the extracted collaborations into a DataFrame for easier manipulation and visualization.

C. Network Construction

1) Graph Initialization: Using NetworkX, we initialized an undirected graph.
2) Adding Nodes and Edges: Nodes represented artists, and edges represented collaborations, with weights indicating the number of collaborative tracks between artists.

III. Basic Characteristics and Visualization

A. Basic Analysis

- **Number of Nodes and Edges**:
  - Number of Artists (nodes): 55
  - Number of Connections (edges): 957
- **Top 10 Nodes by Degree**:
  - De La Ghetto: 50
  - Rauw Alejandro: 49
  - Myke Towers: 48
  - Arcángel: 48
  - Anuel AA: 46
  - Justin Quiles: 46
  - Farruko: 45
  - Bad Bunny: 43
  - Ozuna: 42
  - Nicky Jam: 42
- **Degree Distribution**:

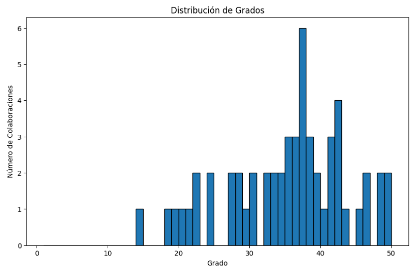

Fig. 1. Distribution

  - Average Degree: 34.84
  - Degree Centralization: Calculated based on node degrees.
- **Density**:
  - Network Density: 0.6415
  - Formula: Density $= \frac{2m}{n(n-1)}$
- **Clustering Coefficient**:
  - Average Clustering Coefficient: Calculated based on the clustering coefficients of all nodes.
- **Connected Components**:
  - Number of Connected Components: 1
  - Largest Connected Component: 55 nodes

B. Additional Centrality Measures

- **Top 10 Artists by Degree Centrality**:
  - De La Ghetto: 0.9259
  - Rauw Alejandro: 0.9074
  - Myke Towers: 0.8889
  - Arcángel: 0.8889
  - Anuel AA: 0.8519
  - Justin Quiles: 0.8519
  - Farruko: 0.8333
  - Bad Bunny: 0.7963
  - Ozuna: 0.7778
  - Nicky Jam: 0.7778
- **Top 10 Artists by Closeness Centrality**:
  - De La Ghetto: 0.9310

- Rauw Alejandro: 0.9153
- Myke Towers: 0.9000
- Arcángel: 0.9000
- Anuel AA: 0.8710
- Justin Quiles: 0.8710
- Farruko: 0.8571
- Bad Bunny: 0.8308
- Ozuna: 0.8182
- Nicky Jam: 0.8182

- **Top 10 Artists by Betweenness Centrality**:
  - De La Ghetto: 0.0174
  - Rauw Alejandro: 0.0164
  - Arcángel: 0.0146
  - Justin Quiles: 0.0146
  - Myke Towers: 0.0142
  - Bad Bunny: 0.0134
  - Lenny Tavárez: 0.0109
  - Sech: 0.0103
  - Farruko: 0.0101
  - Anuel AA: 0.0097
- **Top 10 Artists by Clustering Coefficient**:
  - Almighty: 0.8490
  - Wisin & Yandel: 0.8234
  - Daddy Yankee: 0.8065
  - Jowell & Randy: 0.7882
  - Cosculluela: 0.7866
  - KEVIN ROLDAN: 0.7857
  - El Alfa: 0.7816
  - Beéle: 0.7749
  - Bryant Myers: 0.7718
  - Ryan Castro: 0.7698

C. *Visualization of the Network*

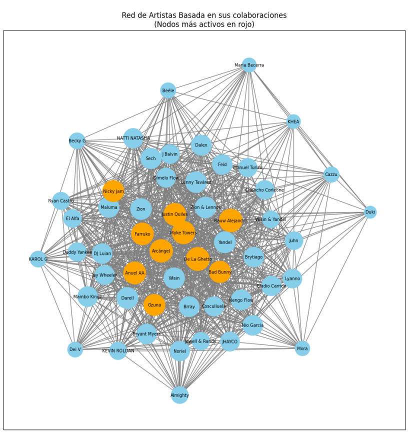

Fig. 2. Spotify network

IV. Link Prediction Analysis

A. *Common Neighbors*

The Common Neighbors metric is based on the premise that two nodes are more likely to form a link if they share a significant number of common neighbors. This intuition comes from the principle of "friends of my friends are my friends," suggesting that nodes with many common neighbors have a higher probability of being connected.

- **Ozuna - Justin Quiles: 39**

The highest result in link prediction based on Common Neighbors is the pair Ozuna - Justin Quiles with a value of 39. This means that Ozuna and Justin Quiles share 39 common neighbors in the collaboration network. This high number of common neighbors suggests a strong probability that these two artists will collaborate in the future if they haven't already. The high number of common neighbors indicates that both artists have extensively collaborated with other artists in the network, creating multiple indirect connections between them.

B. *Jaccard Coefficient*

The Jaccard Coefficient measures the similarity between two nodes by comparing the size of the set of common neighbors to the size of the set of all unique neighbors of both nodes. A high Jaccard Coefficient indicates that two nodes have a high proportion of common neighbors relative to their total neighbors.

- **Ozuna - Justin Quiles: 0.7959**

The highest result in link prediction based on the Jaccard Coefficient is the pair Ozuna - Justin Quiles with a value of 0.7959. This high value suggests that a large proportion of Ozuna's and Justin Quiles' neighbors are common, indicating a strong similarity in their collaboration networks.

C. *Resource Allocation Index*

The Resource Allocation Index assigns more weight to common neighbors with lower degrees. The idea is that two nodes are more likely to connect if they share neighbors who do not have many connections.

- **Ozuna - Justin Quiles: 1.0993**

The highest result in link prediction based on the Resource Allocation Index is the pair Ozuna - Justin Quiles with a value of 1.0993. This value suggests that the likelihood of these two artists collaborating is high because they share common neighbors with fewer connections, reinforcing the bond between them.

D. Preferential Attachment

The Preferential Attachment metric is based on the idea that high-degree nodes tend to form links with other high-degree nodes. The probability that two nodes will connect is proportional to the product of their degrees.

- **Ozuna - Justin Quiles: 1932**

The highest result in link prediction based on Preferential Attachment is the pair Ozuna - Justin Quiles with a value of 1932. This indicates that, since both artists have a high number of connections in the network, the probability of them connecting is very high. This metric highlights the tendency of popular artists to collaborate with each other, further reinforcing their centrality in the network.

V. Conclusions

Based on the link prediction metrics, Ozuna and Justin Quiles have a high likelihood of collaborating in the future. The strong common neighbors, high Jaccard Coefficient, significant Resource Allocation Index, and high Preferential Attachment values all suggest that these two artists are well-connected within the urbano music collaboration network and are likely to form a collaboration soon.

It is important to note that urban artists tend to collaborate frequently, but they do not always repeat collaborations with the same artists. Instead, they try to diversify their collaborations. This behavior contributes to a dynamic and interconnected network. Moreover, attempts to analyze other genres revealed that the urbano genre exhibits the highest frequency of collaborations, making it particularly unique in this regard.

What was your expectation?
My expectation was that artists like Ozuna, Bad Bunny, or Anuel AA would frequently appear in the collaboration predictions due to their prominent roles in the urbano genre. However, I did not anticipate Justin Quiles to emerge so significantly. This result makes sense upon reflection, as many of Justin Quiles' songs feature collaborations with other artists.

How do the results compare to your expectations?
The results partially align with my expectations. While Ozuna frequently appears as expected, the prominence of Justin Quiles was surprising. The metrics consistently identified Ozuna and Justin Quiles as the pair with the highest probability of collaboration, which confirms the robustness of these predictions and highlights the extensive collaborative nature of these artists.

References

[1] Wright, G., Yasar, K. (2022, December 12). Social networking. WhatIs.

[2] Kenton, W. (2022, February 11). Social Networking Service (SNS): characteristics and risks. Investopedia. https://www.investopedia.com/terms/s/social-networking-service-sns.asp

[3] Bello, M. (2023, June 10). Online communities vs social networks: What's the difference? Amity. https://www.amity.co/blog/online-communities-vs-social-networks-whats-the-difference

[4] Disney, A. (2023, August 14). Social network analysis: Understanding centrality measures. Cambridge Intelligence. https://cambridge-intelligence.com/keylines-faqs-social-network-analysis/

[5] Jackson, M. O. (2010). Social and Economic Networks. Princeton University Press.

[6] Newman, M. E. J. (2010). Networks: An Introduction. Oxford University Press.

[7] Barabási, A.-L. (2016). Network Science. Cambridge University Press.

[8] Hanneman, R. A., Riddle, M. (2005). Introduction to Social Network Methods. University of California, Riverside. http://faculty.ucr.edu/ hanneman/

[9] Borgatti, S. P., Everett, M. G., Johnson, J. C. (2018). Analyzing Social Networks. SAGE Publications.

[10] Wasserman, S., Faust, K. (1994). Social Network Analysis: Methods and Applications. Cambridge University Press.

[11] Granovetter, M. S. (1973). The Strength of Weak Ties. American Journal of Sociology, 78(6), 1360-1380.

[12] Scott, J. (2017). Social Network Analysis. SAGE Publications.

[13] Freeman, L. C. (1978). Centrality in social networks conceptual clarification. Social Networks, 1(3), 215-239.

[14] Kadushin, C. (2012). Understanding Social Networks: Theories, Concepts, and Findings. Oxford University Press.

[15] Serrat, O. (2017). Social Network Analysis. In Knowledge Solutions (pp. 39-43). Springer.

[16] Christakis, N. A., Fowler, J. H. (2009). Connected: The Surprising Power of Our Social Networks and How They Shape Our Lives. Little, Brown and Company.

[17] Wellman, B., Berkowitz, S. D. (1988). Social Structures: A Network Approach. Cambridge University Press.