From Bach to Snoop Dogg: Mapping Musical Influence on Spotify

Mark L. Shelton II
Virginia Tech
Blacksburg, Virginia, USA
sheltonml@vt.edu

Charles P. Wiecking
Virginia Tech
Blacksburg, Virginia, USA
wieckingcp23@vt.edu

Abstract

Music streaming platforms such as Spotify have reshaped the music industry by fostering complex networks of artist collaborations. Understanding these networks helps identify influential artists, discover genre-based communities, and improve recommendation strategies. This study examines the Spotify Artist Feature Collaboration Network, consisting of over 150,000 artists and 300,000 collaborations. We address four key questions: (1) Who are the most influential artists based on network position? (2) How are genre-based communities structured? (3) How does network centrality correlate with artist popularity? (4) Can network metrics enhance recommendation strategies? To identify influential artists, we apply centrality measures including degree, eigenvector, betweenness centrality, and PageRank. Results show that classical artists like Johann Sebastian Bach dominate in degree and eigenvector centrality, while modern artists like Snoop Dogg rank high in betweenness centrality, reflecting their roles as genre-bridging collaborators. Using the Louvain method, we detect 1,523 communities, with the largest containing over 24,000 artists spanning genres such as EDM, hip-hop, and pop. Smaller communities often align with niche genres. Our analysis finds a moderate positive correlation (≈ 0.34) between centrality and popularity. Graph-based recommendation methods show that Adamic/Adar and Jaccard favor genre consistency, while Preferential Attachment highlights globally popular artists.

CCS Concepts

• **Computing methodologies → Machine learning**; *Learning latent representations*; • **Information systems → Social networks**; **Recommender systems**; *Graph-based recommendation systems.*

Keywords

Network analysis, Community detection, Artist recommendation, Spotify collaboration network, Centrality metrics, Graph-based recommendation

ACM Reference Format:
Mark L. Shelton II and Charles P. Wiecking. 2025. From Bach to Snoop Dogg: Mapping Musical Influence on Spotify. In *CS 5664 Class Project - Social Media Analytics, Spring 2025, Virginia Tech, Blacksburg, VA, USA.* ACM, New York, NY, USA, 9 pages.

1 Introduction

The music industry is increasingly shaped by patterns of collaboration between artists. Music streaming platforms such as Spotify, Apple Music, and SoundCloud not only enable musicians to reach global audiences but also foster complex collaboration networks that often defy or define traditional genre boundaries. Understanding these networks is crucial for identifying influential figures, uncovering hidden communities, and designing better recommendation strategies.

Analyzing these networks requires robust methods that capture both direct and indirect relationships among artists. Network science techniques are particularly valuable as they reveal both popularity dynamics and genre-based clustering.

In this project, we analyze a large-scale dataset of artist collaborations sourced from Spotify, comprised of over 150,000 artists and 300,000 collaboration edges. We aim to explore the network's structure and dynamics using network science methods and graph-based recommendation algorithms. Our key research questions are: (1) Who are the most influential artists based on their network position? (2) How are genre-based communities structured within the collaboration graph? (3) How does network centrality correlate with artist popularity? (4) Can network metrics enhance artist recommendation strategies?

To address these questions, we apply centrality measures (degree, eigenvector, PageRank, and betweenness) to identify influential artists and use the Louvain method for community detection. We also investigate proximity-based measures like Jaccard's coefficient, Adamic/Adar score, and Preferential Attachment to enhance artist recommendations. Visualization tools such as Gephi support structural interpretation and reveal genre-based clusters.

By integrating classical network science methods with modern recommendation frameworks, this study sheds light on hidden structures within Spotify's artist network. These insights can inform artist discovery, promotion strategies, and personalized music recommendations.

2 Background and Related Works

Analyzing music collaboration networks provides insights into how artists influence each other and how genre-based communities form. Social network analysis techniques, such as centrality measures and community detection, are crucial for understanding these complex structures.

2.1 Music Collaboration Networks Studies

2.1.1 Spotify Artist Collaboration Network. South (2018) conducted a foundational study on the structure of Spotify's artist collaboration network, modeling it as an undirected graph where nodes represented artists and edges denoted collaborations. The study applied centrality measures, including degree and eigenvector centrality, to identify influential artists, revealing a heavy-tailed degree distribution typical of social networks. Additionally, South performed community detection using the Louvain method which uncovered genre-based clusters, whic showed the natural emergence of genre communities within the network. These findings

CS 5664 Class Project, Virginia Tech, Blacksburg, VA, USA
2025.

validate the effectiveness of network analysis for examining music collaboration dynamics. Building on this work, our project incorporates proximity measures and recommendation-based frameworks to further analyze artist influence and predict potential collaborations [10].

2.1.2 Small-World Properties in Spotify Networks. Bush (2025) examined small-world properties in Spotify collaboration networks, focusing on clustering and diameter analysis. The study found that popular artists often act as hubs within dense clusters, reflecting small-world characteristics. Using the Louvain method, Bush identified significant collaboration clusters aligned with genre and regional patterns, showing how both genre and geographical proximity shape collaboration. This study underscores the importance of examining clustering when analyzing collaboration networks. In our analysis, we also use the Louvain method, focusing specifically on genre-based clustering [5].

2.1.3 Popularity and Centrality in Spotify Networks. Mitchell (2020) explored the relationship between popularity and centrality, noting that clusters of high-degree nodes often correspond to popular genres such as rap. The study found that when less popular artists were excluded, rap and hip-hop artists formed dense clusters, while classical artists maintained diverse but less interconnected ties. This work informs our correlation analysis between network centrality and artist popularity, as we also observe a moderate positive relationship between these metrics [11].

2.2 Network Science Applications in Music Recommendation

Previous studies have primarily focused on analyzing collaboration networks through centrality and community detection. However, few have integrated proximity-based recommendations within these networks. Our project addresses this gap by combining centrality analysis, community detection, and link prediction methods (such as Jaccard and Adamic/Adar) to enhance artist recommendations. This approach not only uncovers influential artists and communities but also suggests potential collaborations.

2.3 Summary of Gaps and Our Contribution

While prior research has successfully mapped collaboration structures and examined the relationship between centrality and popularity, they have not sufficiently explored graph-based recommendation methods. Our project integrates both structural and recommendation analysis, using proximity metrics to identify potential collaborations and improve artist discovery. This dual approach bridges the gap between pure network analysis and practical recommendation strategies.

3 Approach

The purpose of this project is to identify which artists from the "Spotify Artist Feature Collaboration Network" are more influential within specific genres. By analyzing the network structure and genre-based clustering, we aim to determine which artists serve as central figures in their genres and examine how an artist's influence correlates with their popularity. Additionally, we leverage network metrics to develop artist recommendation strategies that predict potential future collaborations.

Our approach is structured into eight main steps, which are detailed in the following sections.

3.1 Data Collection and Preprocessing

We obtained the *Spotify Artist Feature Collaboration Network* dataset from Kagglehub. The dataset is available at https://www.kaggle.com/datasets/jfreyberg/spotify-artist-feature-collaboration-network. Once downloaded, we loaded the primary data files containing artists and collaboration information. The `nodes.csv` file contains artist metadata such as name, Spotify ID, popularity, followers, genres, and chart hits. The `edges.csv` file represents collaboration edges between artists, identified by their Spotify IDs. In this dataset, edges correspond to features (featured artists on songs), and nodes represent the artists themselves (musical authors of songs). The edges are undirected and are only stored once. The artist information was scraped from the Spotify API and kworb.net.

After loading the data, we conducted an exploratory analysis to understand its structure and content. This included examining both numerical and text-based fields before constructing the graph. We performed a missing value check to assess nulls in metadata and collaboration edges, as well as sampled genre and chart hit distributions to preview the formatting and diversity of genre and chart metadata. This step provided valuable insights into artist-level features prior to network modeling.

3.2 Network Construction

We modeled the Spotify Artist Collaboration Network as an undirected graph, where nodes represent artists and edges indicate collaborations. Each edge signifies a musical partnership, either through direct features or shared credits.

We built the graph by creating an adjacency list from collaboration pairs extracted from the dataset. Metadata, including artist name, popularity, followers, genres, and chart history, was added to each node. This enriched graph structure serves as the foundation for analyzing centrality, community detection, and recommendation strategies.

3.3 Centrality Analysis

To identify influential artists within the Spotify collaboration network, we compute centrality metrics that quantify the prominence of individual nodes based on their position. Centrality measures are essential in network analysis, as they help identify artists who hold key positions within the collaboration structure.

We use the following four centrality metrics, each providing a unique perspective on influence:

(1) **Degree Centrality:** Measures the number of direct connections an artist has, indicating popularity through direct collaborations. This metric is especially relevant for identifying artists who frequently work with others, acting as central hubs within the network [6]. It is calculated using:

$$C_D(v) = \frac{\deg(v)}{N-1}$$

Here, $\deg(v)$ is the degree of node v, and N is the total number of nodes. High degree centrality reflects direct popularity and widespread influence.

(2) **Eigenvector Centrality:** Extends degree centrality by considering not only the number of connections but also their quality. Artists connected to other influential artists receive higher scores, reflecting indirect influence through association [4]. The eigenvector centrality score is given by:

$$C_E(v) = \frac{1}{\lambda} \sum_{u \in \Gamma(v)} A_{vu} \cdot C_E(u)$$

This measure is crucial for identifying artists embedded within influential clusters, rather than just having many connections.

(3) **Betweenness Centrality:** Quantifies how often an artist appears on the shortest paths between other nodes. This metric identifies artists who act as bridges between communities, facilitating cross-genre collaborations [6]. It is calculated using:

$$C_B(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

Artists with high betweenness centrality are pivotal in connecting otherwise disparate groups within the network.

(4) **PageRank:** Ranks artists by their relative importance, considering both direct connections and the centrality of their neighbors. This approach reveals artists who are not only well-connected but also linked to other prominent artists [9]. The PageRank formula is as follows:

$$PR(v) = \frac{1-d}{N} + d \sum_{u \in \Gamma(v)} \frac{PR(u)}{\deg(u)}$$

The damping factor d, usually set to 0.85, models the probability of continuing a random walk. PageRank is useful for identifying artists embedded in clusters of influence.

By employing these four centrality metrics, we capture various dimensions of network influence: direct connections, association with influential nodes, community bridging, and relative importance. This multifaceted approach helps identify artists who shape collaboration dynamics within the Spotify network.

3.4 Community Detection

The purpose of this section is to identify densely connected groups of artists within the Spotify collaboration network. By detecting structural communities, we aim to understand how musical genres, cultural scenes, or industry structures influence collaboration patterns. Community detection also helps uncover genre-based clusters, providing insight into how artists organize within the collaboration ecosystem.

To detect communities efficiently in this large, undirected graph, we applied the Louvain method. This algorithm optimizes modularity, a measure of the density of edges within communities compared to between communities:

$$Q = \frac{1}{2m} \sum_{ij} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j)$$

In this formula, A_{ij} is the adjacency matrix, k_i and k_j represent the degrees of nodes i and j, m is the number of edges, and δ is an indicator function that equals 1 if nodes i and j belong to the same community. The Louvain method iteratively optimizes this modularity through two phases: local optimization of communities and aggregation of nodes. This method is well-suited for large networks due to its efficiency and scalability. It effectively captures modular structures, making it suitable for analyzing genre-based clusters within the Spotify artist collaboration network [3].

After applying the Louvain method, we analyzed the resulting community structure by counting the number of detected communities and examining the largest by size. We then sampled artists from the most prominent communities to understand their genre alignment. This analysis addresses the goal of detecting and analyzing communities within the Spotify network and supports the interpretation of network modularity in terms of genre and collaboration behavior.

3.5 Correlation Analysis

To assess the relationship between network influence and commercial success, we calculate the Pearson correlation coefficient between centrality measures (degree and eigenvector centrality) and artist popularity. The Pearson correlation coefficient is given by:

$$\rho = \frac{\sum (x_i - \mu_x)(y_i - \mu_y)}{\sqrt{\sum (x_i - \mu_x)^2 \sum (y_i - \mu_y)^2}}$$

A positive coefficient indicates that higher centrality generally aligns with greater popularity, consistent with previous findings on network influence and commercial success [8].

3.6 Visualization

The purpose of this section is to generate visual summaries of the collaboration graph to support structural interpretation. Visualizing the network structure helps identify key patterns, influential artists, and genre-based clusters. We use a series of visualization techniques to show the graph's structural characteristics.

We utilize the following visualization methods to capture different aspects of the network:

(1) **Degree Distribution (Log-Log Scale):** This plot assesses whether the network exhibits scale-free properties, which are typical in social and collaboration networks. It reveals the presence of a few highly connected hub artists compared to the majority with fewer connections. Identifying these hubs helps reveal central figures within the network.

(2) **Community Structure Visualization:** To illustrate genre-based clustering, we generate a *Sampled Subgraph Colored by Community* using a 200-node sample from the largest connected component (Figure 1). Nodes are colored according to their Louvain community ID, providing a visual snapshot of how different communities cluster together. This helps in identifying cohesive genre-based groups.

(3) **Community Size Distribution:** This plot demonstrates the variation in community sizes, helping determine whether the network is dominated by a few large clusters or consists

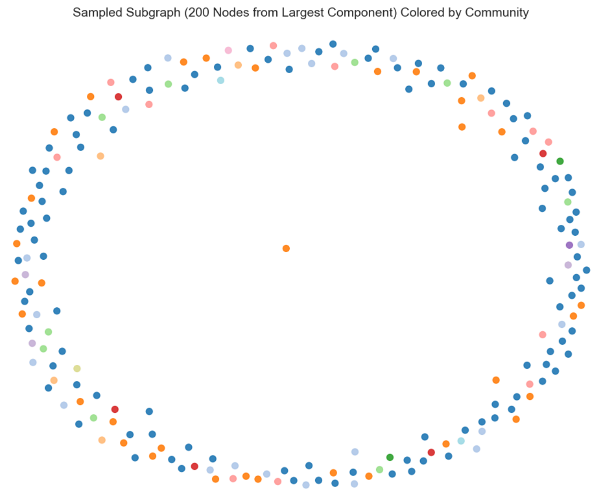

Figure 1: Sampled Subgraph (200 Nodes) using `networkx`

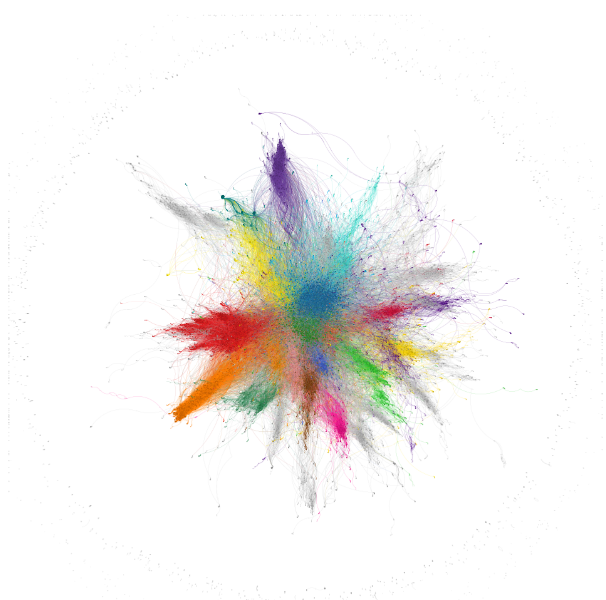

Figure 2: Full Network Colored by Louvain Community using Gephi

of many smaller groups. Understanding the size distribution aids in interpreting the network's modularity.

(4) **Connected Component Sizes:** To analyze network fragmentation, we examine the size distribution of connected components. This visualization reveals whether the network is mostly connected or contains isolated subgraphs, indicating the extent of artist collaboration across the network.
(5) **Interactive Graph Export:** To facilitate further analysis, we export the complete graph to Gephi in the `.gexf` format (Figure 2). This export enables advanced layout adjustments and modularity-based visualization, allowing for dynamic exploration of the network structure.

By combining these visualizations, we gain a comprehensive understanding of the Spotify artist collaboration network. These methods reveal key structural features, identify influential artists, and potentially reveal genre-based clusters, offering valuable insights into collaboration dynamics.

3.7 Artist Recommendation Modeling

The purpose of this section is to develop graph-based methods to recommend potential artist collaborations or similar artists. We employ unsupervised link prediction techniques that leverage structural proximity within the collaboration graph. To ensure fair comparisons, we preprocess the data by filtering artists based on degree, metadata presence, and community membership.

To select representative artists for the recommendation system, we applied the following filtering criteria:

- Degree between 30 and 1000, to avoid overly sparse or dominant hub nodes.
- Known Louvain community label, to maintain genre context.
- Valid artist name, to ensure accurate and interpretable recommendations.

From the filtered set, we selected three representative artists from distinct musical backgrounds: Snoop Dogg (Rap), John Williams (Film Score), and Pitbull (Pop/Latin). These artists were chosen because they meet the connectivity criteria, belong to different Louvain communities to ensure diversity, and have complete metadata. This selection provides a balanced evaluation of recommendation methods across varied musical contexts.

We implement the following three graph-based recommendation methods, each capturing a different aspect of structural similarity:

(1) **Adamic/Adar Similarity:** This method prioritizes artists with rare shared collaborators, giving higher scores to pairs connected through low-degree nodes. It is particularly useful for discovering niche collaborations where artists share a few but significant connections [1].

$$\text{Adamic/Adar}(x, y) = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{\log(\deg(z))}$$

This metric is beneficial for identifying collaborations that are not immediately obvious but significant within specific genres or sub-communities.

(2) **Jaccard Similarity:** This method measures the ratio of shared to total collaborators, emphasizing the proportion of mutual connections. It is well-suited for recommending artists who frequently collaborate within the same community [7].

$$J(x, y) = \frac{|\Gamma(x) \cap \Gamma(y)|}{|\Gamma(x) \cup \Gamma(y)|}$$

The Jaccard coefficient is ideal for maintaining genre consistency when suggesting collaborations, as it shows artist pairs with significant overlap in their collaboration networks.

(3) **Preferential Attachment:** This method predicts collaborations based on the product of the degree of two nodes. It

favors high-degree nodes, making it useful for identifying potential collaborations with widely popular artists [2].

$$\mathrm{PA}(x,y) = \deg(x) \times \deg(y)$$

This technique is effective for recommending globally popular artists who are likely to collaborate with other high-degree nodes, irrespective of genre alignment.

By using these three methods, we capture diverse perspectives on collaboration potential: Adamic/Adar focuses on niche and rare collaborations, Jaccard ensures genre consistency, and Preferential Attachment identifies prominent artists likely to form new connections. Combining these techniques provides a balanced approach to artist recommendation within the Spotify collaboration network.

3.8 Comparative Evaluation

The purpose of this section is to assess how different structural recommendation strategies perform when applied to artists from diverse musical backgrounds. We evaluate the effectiveness of each method by comparing the recommendations for three representative artists: Snoop Dogg (Rap), John Williams (Film Score), and Pitbull (Pop/Latin). These artists were chosen for their distinct musical genres and collaboration patterns by applying the three methods: Adamic/Adar, Jaccard, and Preferential Attachment. This comparative analysis deepens our understanding of how structural features influence recommendation quality and demonstrates how method effectiveness varies by artist type and community structure. The evaluation also addresses the project's goal of improving artist recommendations based on network topology.

4 Experiment

4.1 Introduction

In this experiment, we analyze the Spotify Artist Feature Collaboration Network to address key research questions related to artist influence, community structure, centrality-popularity correlation, and recommendation strategies. The dataset consists of nodes representing artists with metadata such as name, popularity, followers, and genres, while edges denote collaborations between artists identified by their Spotify IDs.

We aim to answer the following research questions:

(1) Who are the most influential artists based on their network position?
(2) How are genre-based communities structured within the collaboration graph?
(3) How does network centrality correlate with artist popularity?
(4) Can network metrics be leveraged to develop artist recommendation strategies?

After data preprocessing and graph construction, we apply centrality analysis to identify influential artists and use community detection to uncover genre-based clusters. We also examine the correlation between network position and popularity and develop graph-based methods for artist recommendation. The results provide insights into collaboration dynamics and support improved artist discovery.

4.2 Basic Graph Information

To provide context for the following experiments, we present basic structural information about the Spotify Artist Collaboration Network graph. The graph is constructed as an undirected network where nodes represent artists and edges indicate collaborations.

After loading the dataset and constructing the graph, we obtained a total of 153,327 nodes (artists) and 300,386 edges (collaborations). The network has an average degree of approximately 3.92, indicating that most artists have a small number of direct collaborations, while a few central figures have significantly more.

To understand the distribution of popularity and followers among artists, we performed a basic statistical analysis using the `.describe()` function. The summary statistics are as follows:

- **Follower Count:** The mean follower count is approximately 86,224, with a standard deviation of 940,100. The median follower count is 362, showing a highly skewed distribution where a few artists have exceptionally large followings. The maximum follower count exceeds 102 million, while 25% of artists have fewer than 24 followers.
- **Popularity Score:** The mean popularity score is 21.16, with a standard deviation of 18.34. The median score is 18, indicating that most artists are relatively less popular. The maximum popularity score is 100, with 25% of artists having a score of 4 or less.

Understanding the basic structure is essential for subsequent analysis, as it reveals the underlying connectivity patterns that influence centrality and community formation. Community detection using the Louvain method revealed 1,523 distinct communities, with the largest containing over 24,000 artists. These communities often align with musical genres or cultural connections, indicating that collaboration frequently occurs within stylistically similar groups.

To further illustrate the degree distribution of the collaboration network, we generated a log-log plot (Figure 3) which shows a heavy-tailed distribution, consistent with the scale-free behavior typical of social and collaborative networks. Additionally, we plotted the distribution of connected component sizes (Figure 4) to show network fragmentation.

This basic graph information forms the foundation for the following experiments, guiding our analysis of artist influence, community structure, popularity correlations, and recommendation strategies.

4.3 Research Question 1: Identifying Influential Artists

The purpose of this experiment is to analyze the structural properties of the collaboration graph to identify influential artists. After constructing the graph with 153,327 nodes and 300,386 edges, we enriched the artist nodes with metadata, including name, popularity, followers, genres, and chart history. This undirected graph enabled a comprehensive network analysis.

To identify influential artists, we computed four centrality metrics: degree centrality, eigenvector centrality, betweenness centrality, and PageRank. These metrics measure direct collaborations, influence through connections, strategic positioning, and embeddedness within influential clusters, respectively.

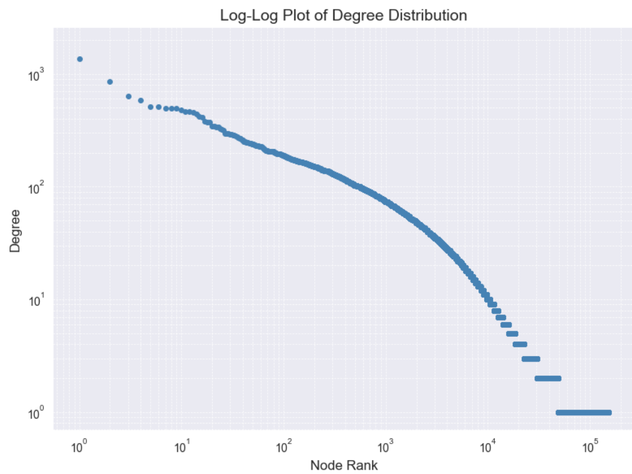

Figure 3: Log-Log Plot of Degree Distribution

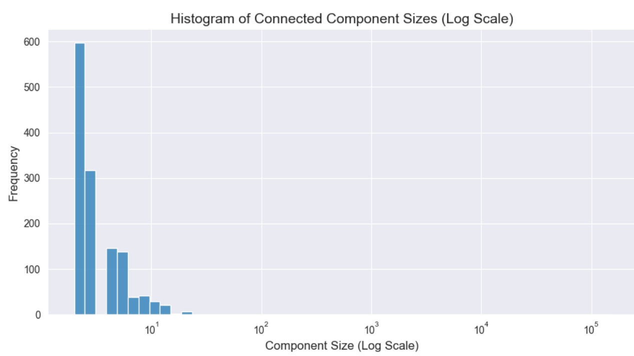

Figure 4: Histogram of Connected Component Sizes (Log Scale)

Our analysis revealed that classical and traditional artists, such as Johann Sebastian Bach and Traditional, consistently rank high in both degree and eigenvector centrality, indicating their roles as central hubs with numerous collaborations. In contrast, betweenness centrality and PageRank identified modern and versatile artists like Snoop Dogg, R3HAB, and Diplo, reflecting their roles as genre-bridging collaborators. These patterns demonstrate how traditional compositions are widely referenced, while modern artists facilitate cross-genre interactions.

To illustrate the most influential artists by each centrality metric, we present the top five artists identified by degree, eigenvector, and betweenness centrality:

These results demonstrate that classical artists, like Johann Sebastian Bach, hold structural influence due to their numerous and well-connected collaborations. In contrast, modern artists like Snoop Dogg and Diplo exhibit influence through cross-genre connections, reflecting their roles as community bridges.

The distribution of eigenvector centrality values, shown in Figure 6, shows the concentration of influence among a few central artists. Similarly, the PageRank distribution in Figure 5 illustrates how a small set of highly influential nodes dominates the network.

Artist	Genre(s)	Centrality
J. S. Bach	Baroque, Classical	0.01162
Traditional	Puirt-a-Beul, Sleep	0.00894
Mc Gw	Funk Carioca, Funk MTG	0.00560
MC MN	Funk Carioca, Funk MTG	0.00412
Jean Sibelius	Classical, Impressionism	0.00378

Table 1: Top 5 Artists by Degree Centrality

Artist	Genre(s)	Centrality
J. S. Bach	Baroque, Classical	0.29813
Traditional	Puirt-a-Beul, Sleep	0.22954
Mc Gw	Funk Carioca, Funk MTG	0.14371
MC MN	Funk Carioca, Funk MTG	0.10590
Jean Sibelius	Classical, Impressionism	0.09720

Table 2: Top 5 Artists by Eigenvector Centrality

Artist	Genre(s)	Centrality
Traditional	Puirt-a-Beul, Sleep	0.03537
Snoop Dogg	Gangster Rap, Hip Hop	0.03461
R3HAB	Dance Pop, EDM	0.03146
J. S. Bach	Baroque, Classical	0.02615
Diplo	EDM, Electro House	0.02347

Table 3: Top 5 Artists by Betweenness Centrality (Sampled)

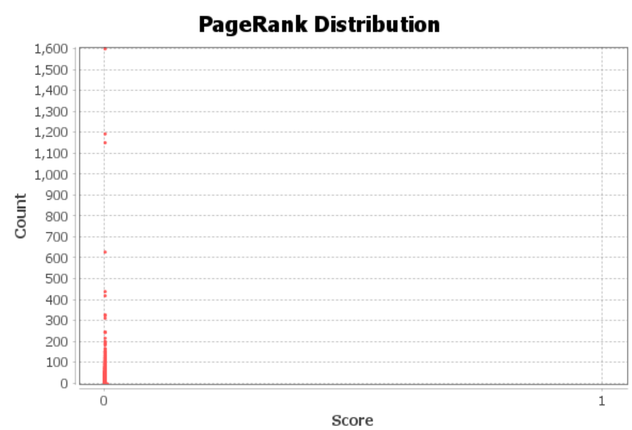

Figure 5: Gephi-Generated PageRank Distribution

4.4 Research Question 2: Community Structure and Genres

To identify genre-based clusters within the Spotify collaboration network, we applied the Louvain community detection algorithm. The method revealed 1,523 distinct communities, indicating a modular structure where artists naturally group according to genres or cultural similarities.

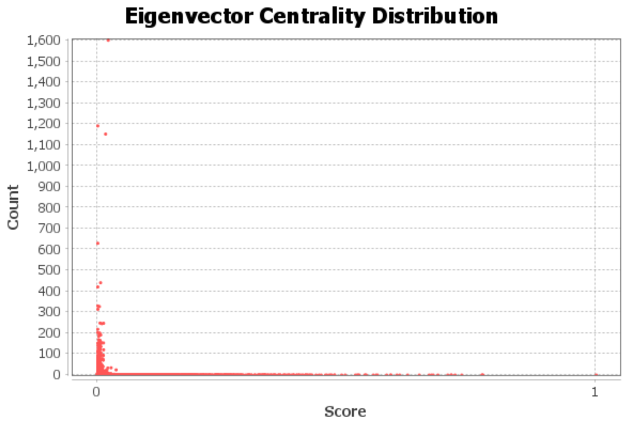

Figure 6: Gephi-Generated Eigenvector Centrality Distribution

The largest detected community consists of over 24,000 artists, encompassing interconnected genres such as EDM, hip hop, and pop. In contrast, smaller communities often correspond to niche genres, showing how artists with similar styles cluster together. For example, one community prominently featured classical composers, while another grouped Latin pop artists, reflecting how collaboration networks mirror genre affiliations.

We visualized the community structure using Gephi (Figure 2), where nodes are colored according to their Louvain community assignments. The color-coded graph shows dense intra-community connections and sparse inter-community links, illustrating how genre-based clusters form distinct subgraphs. For instance, Community 0 (Blue) represents the largest group, often comprising EDM and pop genres. In contrast, smaller clusters like Community 5 (Pink) and Community 33 (Teal) likely correspond to niche genres or regional collaborations.

Among the top 10 largest communities, the Blue community (Community 0) contains 24,387 artists (15.9% of the total network), followed by the Red community (Community 4) with 13,652 artists (8.9%) and the Green community (Community 11) with 8,701 artists (5.7%). These larger communities typically feature genres with broad appeal and diverse collaborations. In contrast, smaller communities such as the Royal Blue community (Community 3) with 4,132 artists (2.7%) and the Deep Pink community (Community 22) with 3,560 artists (2.3%) are more tightly knit, reflecting specialized genres or cultural scenes.

The modular nature of the network suggests that collaborations are generally genre-consistent, with some artists acting as bridges between genres, connecting multiple communities. Identifying these bridge artists is crucial as they play a key role in blending genres and expanding the reach of stylistically diverse music. Understanding these structural patterns can help predict collaboration trends and inform targeted recommendations.

4.5 Research Question 3: Correlation between Centrality and Popularity

To examine the relationship between network influence and commercial success, we computed the Pearson correlation coefficient between centrality metrics (degree and eigenvector centrality) and artist popularity. The Pearson correlation coefficient is appropriate for this analysis as it quantifies the linear relationship between numerical variables, allowing us to assess how centrality relates to popularity.

The analysis revealed a moderate positive correlation of ≈ 0.34 for both metrics, indicating that well-connected or influential artists tend to be more popular. This positive correlation suggests that while higher centrality generally aligns with greater popularity, the relationship is not exceptionally strong.

However, the results also indicate that centrality is not the sole determinant of an artist's prominence within the network. External factors such as marketing, genre, and cultural relevance significantly influence popularity, suggesting that network position alone does not guarantee commercial success.

These findings imply that while centrality contributes to an artist's visibility and influence within the collaboration network, external factors play a substantial role in shaping their popularity. Understanding this nuanced relationship is essential for developing more accurate prediction models for artist success.

4.6 Research Question 4: Recommendation Strategies

To develop artist recommendation strategies, we applied three graph-based methods: Adamic/Adar, Jaccard, and Preferential Attachment. Each method uses network topology to identify potential collaborations by analyzing structural similarity and connectivity patterns.

(1) **Adamic/Adar and Jaccard Methods:** These methods prioritize shared neighbors and direct collaborations, producing recommendations that align with the target artist's genre. They are particularly effective for maintaining genre consistency when suggesting potential collaborations.
(2) **Preferential Attachment Method:** This method often suggests highly popular artists regardless of genre, reflecting a bias toward nodes with many connections. It is useful when the goal is to recommend globally recognized artists or those with extensive collaboration networks.

Choosing the appropriate method depends on the recommendation context: whether the aim is to preserve genre alignment or to emphasize broad influence through high-degree nodes.

Evaluation: We assessed the effectiveness of these methods by performing a three-way comparison using three representative artists from distinct musical backgrounds: Snoop Dogg (Rap), John Williams (Film Score), and Pitbull (Pop/Latin).

- **Snoop Dogg:** Adamic/Adar and Jaccard generated genre-consistent recommendations, including artists like 2 Chainz, Diddy, and Big Sean. Preferential Attachment, however, recommended more globally popular artists such as Steve Aoki and Diplo.
- **John Williams:** Both Adamic/Adar and Jaccard recommended classical and film score composers, including Jean Sibelius, Andrea Bocelli, and Hans Zimmer. Preferential Attachment suggested high-degree artists, like the Iceland Symphony Orchestra and Vera Lynn.

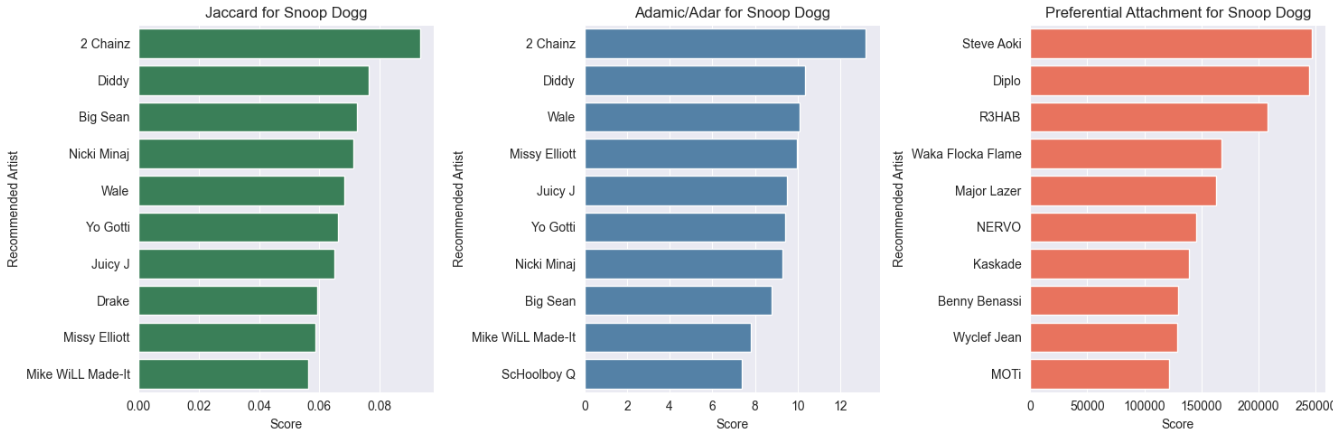

Figure 7: Comparison of recommendation methods for Snoop Dogg (Rap)

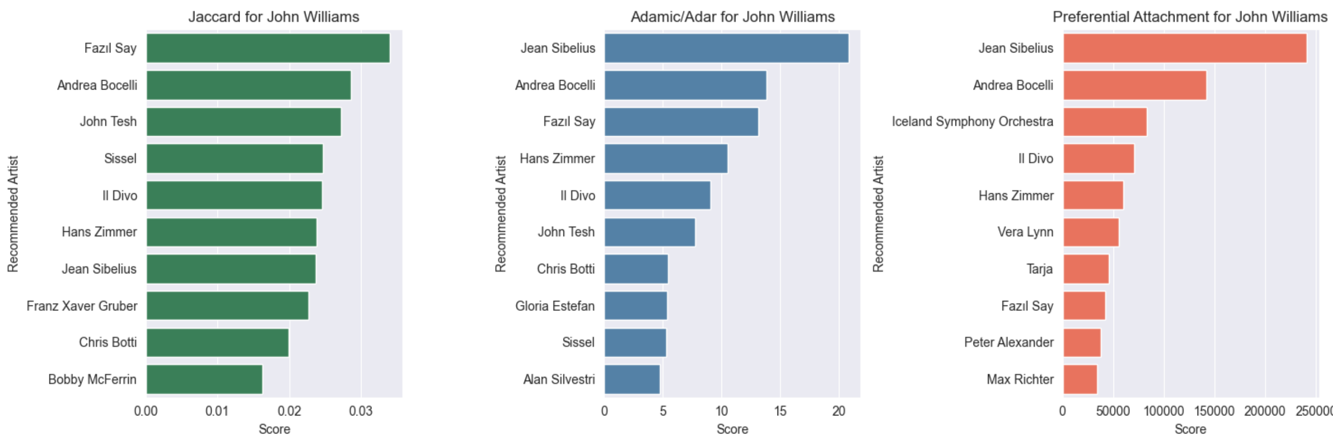

Figure 8: Comparison of recommendation methods for John Williams (Film Score)

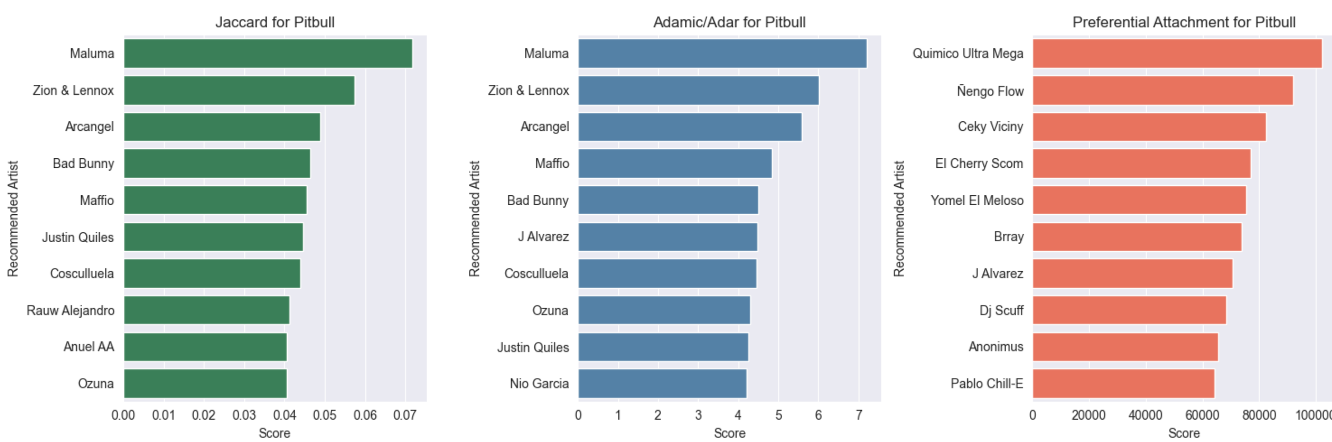

Figure 9: Comparison of recommendation methods for Ptibull (Pop/Latin)

- **Pitbull:** Adamic/Adar and Jaccard produced genre-aligned recommendations from Latin and reggaeton artists, including Maluma, Zion & Lennox, and Bad Bunny. Preferential Attachment favored high-degree collaborators such as Quimico Ultra Mega and Ñengo Flow.

Figures 7, 8, and 9 illustrate the comparison between methods for the three target artists. Each chart shows the top recommendations from each method, demonstrating that Adamic/Adar and Jaccard maintained genre consistency, while Preferential Attachment tended to favor globally popular artists.

These results indicate that Adamic/Adar and Jaccard are more suitable for genre-focused recommendations, while Preferential Attachment is more effective for promoting artists with extensive collaborations or global reach. The choice of method should align with the specific recommendation goal: maintaining genre cohesion or maximizing exposure through influential artists.

4.7 Discussion

Our experiments reveal that the Spotify Collaboration Network exhibits a modular and community-driven structure, where influential artists generally have high centrality scores. However, while centrality metrics provide insights into network influence, our analysis shows that they do not solely determine an artist's popularity. External factors such as marketing, genre appeal, and cultural relevance play significant roles, indicating that network position alone is insufficient for predicting commercial success.

Recommendation modeling demonstrated that Adamic/Adar and Jaccard methods consistently produce genre-aligned recommendations, making them suitable for applications where genre consistency is important. In contrast, Preferential Attachment tends to recommend popular, high-degree artists, making it more effective when the goal is to maximize visibility or connect with prominent figures. These differences highlight the need to choose recommendation methods based on the desired balance between maintaining genre integrity and promoting artists with a wide reach.

An important consideration is the trade-off between genre specificity and global popularity. While Adamic/Adar and Jaccard are effective for identifying collaborators within the same genre, they may overlook high-profile opportunities that Preferential Attachment naturally identifies. Conversely, Preferential Attachment may compromise genre consistency by favoring widely popular artists, potentially diluting genre-based recommendations.

Future improvements could involve hybrid approaches that combine the strengths of these methods, integrating genre consistency with visibility to create balanced recommendations. Additionally, incorporating external data on cultural impact or listener engagement could enhance the prediction accuracy, reflecting a more holistic view of artistic influence.

By understanding the strengths and limitations of each method, recommendation systems can be tailored to fit specific contexts, whether aiming for genre-centric collaboration suggestions or broader exposure within the music industry.

5 Conclusion

This project analyzed the structural patterns and recommendation strategies within the Spotify Artist Feature Collaboration Network, comprising over 150,000 artists and 300,000 collaboration edges. The primary objectives were to identify influential artists, detect genre-based communities, examine the relationship between network influence and popularity, and develop graph-based artist recommendations.

We identified key artists and structural hubs using centrality measures such as degree, eigenvector, and sampled betweenness centrality. Community detection with the Louvain algorithm revealed over 1,500 modular clusters, many aligning with genre patterns. Network visualizations further demonstrated the network's modularity and community structures.

We developed recommendation strategies using the Adamic/Adar Index, Jaccard Coefficient, and Preferential Attachment. Adamic/Adar and Jaccard methods provided genre-consistent recommendations, while Preferential Attachment favored globally popular artists. This difference underscores the importance of choosing methods based on the intended balance between genre coherence and visibility.

Our analysis also revealed a moderate positive correlation (≈ 0.34) between centrality and popularity, indicating that influential artists tend to be more popular, though external factors also play a significant role.

These findings contribute to understanding how network structure shapes artistic influence and collaboration. Future work could explore hybrid recommendation models and incorporate listener engagement data to improve accuracy.

References

[1] Lada A. Adamic and Eytan Adar. 2003. Friends and neighbors on the web. In *Proceedings of the International Conference on Research and Development in Information Retrieval (SIGIR)*. 211–218. doi:10.1145/860435.860460

[2] Albert-László Barabási and Réka Albert. 1999. Emergence of scaling in random networks. *Science* 286, 5439 (1999), 509–512. doi:10.1126/science.286.5439.509

[3] Vincent D. Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. 2008. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* 2008, 10 (2008), P10008. doi:10.1088/1742-5468/2008/10/P10008

[4] Phillip Bonacich. 2007. Some unique properties of eigenvector centrality. *Social Networks* 29, 4 (2007), 555–564. doi:10.1016/j.socnet.2007.04.002

[5] Raquel Ana Bush. 2025. Analysis of a Spotify Collaboration Network for Small-World Properties. https://arxiv.org/pdf/2503.09526 Department of Mathematics, University of Massachusetts Dartmouth, North Dartmouth, Massachusetts.

[6] Linton C. Freeman. 1979. Centrality in social networks: Conceptual clarification. *Social Networks* 1, 3 (1979), 215–239. doi:10.1016/0378-8733(78)90021-7

[7] Paul Jaccard. 1901. Étude comparative de la distribution florale dans une portion des Alpes et des Jura. *Bulletin de la Société Vaudoise des Sciences Naturelles* 37 (1901), 547–579.

[8] Javad Mohamadichamgavi, Mahdi Hajihashemi, and Keivan Aghababaei Samani. 2024. An Analysis of Correlation and Comparisons Between Centrality Measures in Network Models. *Journal of Social Structure* 25, 1 (2024), 1–15. doi:10.21307/joss-2024-001

[9] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. 1999. The PageRank Citation Ranking: Bringing Order to the Web. *Stanford InfoLab* (1999). http://ilpubs.stanford.edu:8090/422/ Technical Report.

[10] Tobin South. 2018. Spotify Collaboration Graph Analysis. https://vrs.amsi.org.au/wp-content/uploads/sites/84/2018/04/tobin_south_vrs-report.pdf Undergraduate Vacation Research Scholarship Report, AMSI.

[11] Matthew Roughan Tobin South and Lewis Mitchell. 2020. Popularity and centrality in Spotify networks: critical transitions in eigenvector centrality. https://www.researchgate.net/publication/350004417_Popularity_and_centrality_in_Spotify_networks_critical_transitions_in_eigenvector_centrality ARC Centre ofExcellence for Mathematical Statistical Frontiers (ACEMS).

Received 10 May 2025