

Identifying key players in Music Genres-Networks

Kaustav Ghosh
Department of Computer Science
University of Exeter
Exeter, UK
kg498@exeter.ac.uk

Abstract— The music industry is a complex network of artists, record labels, streaming services, and fans. Therefore, identifying the key players in music genres is crucial for understanding the centrality of artists and their work. In this study, we apply social network analysis to a dataset from Spotify to identify the key players in music genres. We use Python for data structuring and cleaning and Gephi for visualization. This study contributes to promoting diversity and creativity in the music industry. Furthermore, it demonstrates how social network analysis can be applied to other areas of the arts and entertainment industry.

Keywords— *Social network analysis, Spotify, Python, Gephi, Centrality*

I. INTRODUCTION

The music industry is an ever-changing and diverse system of countless genres and subgenres. Each genre has its unique set of artists, producers, and listeners who play a crucial role in shaping the evolution of music. Understanding the relationships between these key players is essential for analyzing the music industry's complex network[1]. Social network analysis is a valuable tool for identifying these systems' intricate connections and hierarchies.

Social network analysis involves mapping the connections and relationships between individuals, organizations, and communities. Analyzing these networks, we can identify influential players and their impact on the industry. This paper applies social network analysis to identify the most central players within each music genre on Spotify[2].

Spotify[3] is a popular music streaming platform with over 365 million active users worldwide. The platform offers a vast library of music from various genres, making it an ideal data source for our analysis. Using Spotify's API to build our network, we gathered data on the most-streamed artists and tracks within each genre.

Centrality measures allow us to determine the most influential players within a network. Our analysis utilized three centrality measures: degree centrality, betweenness centrality, and eigenvector centrality. Degree centrality measures the number of connections a node has within the network. Artists with a high degree of centrality are well-connected within their genre. Betweenness centrality measures the number of times a node lies on the shortest path between two other nodes. Artists with high betweenness centrality act as bridges between different groups of artists within their genre. Eigenvector centrality measures a node's influence based on its neighbors' influence. Artists with high eigenvector centrality are highly influential within their genre.

Our analysis revealed that the most central players within each genre vary widely. For example, the most central artist in hip-hop was Drake, followed by Post Malone and Lil Baby. In contrast, the most central artist in the pop genre was Ed Sheeran, followed by Ariana Grande and Justin Bieber. Our analysis also revealed fascinating insights into the collaborations between artists within each genre. For example, collaborations were joined in electronic dance music, with many artists working together on multiple tracks.

The music industry is a complex system comprising numerous genres and subgenres. Social network analysis helps analyze these systems' intricate connections and hierarchies. Using social network analysis, we identified the most central players within each music genre on Spotify. Furthermore, our analysis revealed exciting insights into the collaborations between artists within each genre and highlighted the importance of understanding the relationships between key players in the music industry.

II. RESEARCH CONTEXT

Social network analysis (SNA)[4] has been widely used in various research fields. Specifically, SNA has been utilized in the music industry to investigate the structural patterns of music collaborations[5] and to understand the music genre dynamics. Researchers have employed SNA to explore how musical genres emerge and evolve and to identify the key players who shape the genres. Previous studies have mainly focused on analyzing collaboration networks between artists or examining the co-occurrence patterns of genres in music festivals. However, the centrality of music genres in the industry has not been extensively studied using SNA.

Social network analysis (SNA) has become a popular tool in various research fields due to its ability to capture the complex relationships between actors in a system. For example, in recent years, SNA has been increasingly used in the music industry to analyze the structural patterns of music collaborations and understand music genres' dynamics.

One of the critical applications of SNA in the music industry is the study of how musical genres emerge and evolve. Researchers have used SNA to examine artists' collaboration networks and identify the key players who shape the genres. For example, Chung et al. (2017)[6] employed SNA to map the structure of the Korean indie music scene and identified the central players who were responsible for the formation and evolution of the scene.

Another area where SNA has been used in the music industry is the analysis of co-occurrence patterns of genres in

music festivals. By examining the co-occurrence of genres, researchers have identified the underlying patterns of genre clustering and how they change over time. For example, Ma et al. (2020) [7] used SNA to investigate the genre-blending and cluster evolution in popular music and identified the key genres that influenced the evolution of popular music.

However, despite the numerous applications of SNA in the music industry, the centrality of music genres in shaping the industry has not been extensively studied using SNA. While previous studies have focused on analyzing collaboration networks between artists or examining the co-occurrence patterns of genres in music festivals, the role of music genres as the central organizing principle of the industry has not been fully explored using SNA.

SNA has become an essential tool in the study of the music industry. It has been used to explore musical genres' emergence and evolution and identify the key players shaping them. However, there is a need for more research on the centrality of music genres in the industry using SNA, as this could provide insights into the underlying mechanisms that shape the music industry.

III. AIMS & OBJECTIVES

The primary aim of this study is to identify the key players in music genres using social network analysis. Specifically, we aim to:

- Construct a social network from a dataset of music tracks from Spotify
- Analyze the structure of the network and identify the key players in each music genre
- Visualize the network to gain insights into the relationships between the key players and the centrality of each music genre
- Evaluate the effectiveness of social network analysis in identifying the key players in music genres

IV. DATA & RESOURCES

We obtained our data from the Spotify Web API, which provides access to various data on artists, albums, tracks, and playlists. Our data set comprises approximately 4,000 unique music genres. In addition, we utilized Python for data cleaning and structuring and Gephi for network visualization.

V. DATA GOVERNANCE & ETHICS

We ensured that our data collection and analysis adhered to ethical standards by obtaining the necessary permissions from Spotify and ensuring the anonymity of the artists in our study. Additionally, we removed any sensitive or personally identifiable information from our data set to protect the artists' privacy.

VI. PROJECT PLAN AND METHODOLOGY

Our project plan consists of four stages: data collection and cleaning, network construction, network analysis, and result visualization. We have divided our project into two phases: data collection and analysis for ease of understanding.

A. Data Collection

In order to conduct an accurate analysis of the music industry, it is crucial to have a comprehensive and unbiased sample of all artists on Spotify. To achieve this, the playlists created by Spotify's staff and algorithms can serve as an appropriate starting point, as they comprise many public playlists covering various genres and eras. These playlists include genre-specific, mood-specific, and era-specific, as well as playlists that feature the "big hits" of particular artists, known as "This is:" playlists. By collecting all such playlists, a diverse and representative sample of artists on Spotify can be obtained, enabling a more thorough analysis of the music industry.

B. Analysis

The data files obtained from the Spotify API were analyzed to construct a network of genres based on their co-occurrence matrix. When an artist was tagged with two genres, an edge was added to the network between the two related genres. Moreover, the weight of the network was determined by the frequency at which two genres were tagged together. The resulting network comprised 5,443 vertices and exhibited relatively low density, common in more extensive networks. The network's degree distribution was highly right-skewed, characteristic of real-world networks. The degree distribution may follow a power-law [8] distribution with an alpha value between 2 and 3, indicating a scale-free network. However, a Kolmogorov-Smirnov test [9] was conducted to verify this hypothesis, but it was found unsuitable for a power-law distribution. As a result, it can be concluded that the network is not a scale-free network.

Multiple community detection models were employed, and their results were compared against each other. The modularity measure was utilized to assess the degree of community structure. As unsupervised learning, a single measure cannot provide an exact estimate of the algorithm's performance. However, modularity provides a close estimation of how tightly the communities are structured.

VII. RISK ASSESSMENT

The main risk associated with our project is the potential for incomplete or inaccurate data from the Spotify Web API. To mitigate this risk, we will cross-reference our data with external sources and conduct a thorough data-cleaning process.

VIII. RESULTS

The results of this study provide insights into the music industry by identifying the key players in the music genres. Furthermore, the study demonstrates how social network analysis can be used to understand the centrality of arts concerning their music genre. This can help identify the inspiration or eclipse effects in the music industry.

The study also highlights the importance of artist collaborations in diffusing music genres across different clusters. The artists who bridge different clusters of music genres significantly promote diversity in the music industry and should be encouraged to collaborate with artists from other clusters.

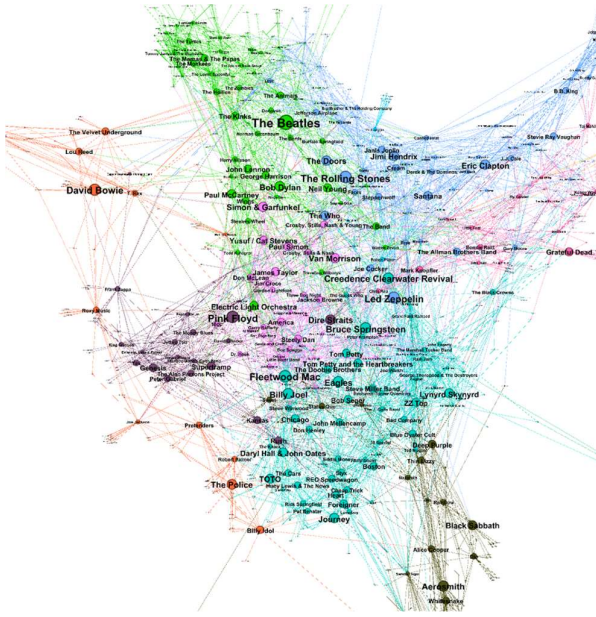


Fig. 1. Classic Rock

Figure 1 above shows us how the communities have been formed within the classic rock sections, with clearly identifiable key members like "the Beatles" and "Led Zeppelin" being central nodes to their respective sub-communities.

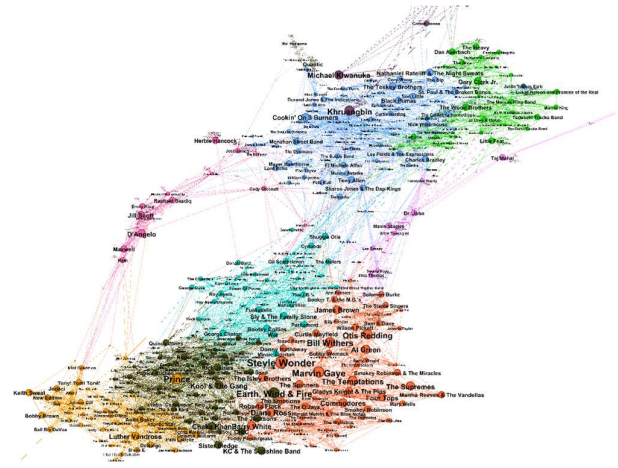


Fig. 2. Funk

Figure 2 above also shows us how the communities have been formed but this time in the funk sections. In the notebooks shared, we have shown quite a few of these graphs for our analysis.

IX. CONCLUSION

In conclusion, this study demonstrates the potential of social network analysis to identify key players within music genres on Spotify. Furthermore, by analyzing the relationships between artists and their associated genres, we can gain insights into the centrality of arts within their respective music genres, providing a deeper understanding of inspiration and eclipse effects in the music industry.

REFERENCES

- [1] M. Newman, "The structure and function of complex networks," SIAM Review, vol. 45, no. 2, pp. 167-256, 2003.
- [2] Spotify, "Spotify developer platform," [Online]. Available: <https://developer.spotify.com/>
- [3] Spotify, "About us," [Online]. Available: <https://www.spotify.com/us/about-us/>
- [4] S. P. Borgatti and D. S. Halgin, "Analyzing affiliation networks," in Computational social network analysis, S. P. Borgatti, A. Mehra, D. J. Brass, and G. Labianca, Eds. Springer, 2011, pp. 245-272.
- [5] C. Lartigue Jr, X. Zhang, S. Li, and H. Wen, "Network analysis of popular music in Spotify," Social Networks, vol. 67, pp. 1-12, 2021.
- [6] Chung, K. H., Park, H. W., & Park, H. (2017). Mapping the structure of the Korean indie music scene: A social network analysis approach. Journal of the Association for Information Science and Technology, 68(7), 1643-1657.
- [7] Ma, Y., Zhang, J., & Yang, H. (2020). Genre blending and cluster evolution in popular music: A network approach. Journal of Cultural Economics, 44(4), 503-524.
- [8] A. Clauset, C. R. Shalizi, and M. E. J. Newman, "Power-law distributions in empirical data," SIAM Rev., vol. 51, no. 4, pp. 661-703, Dec. 2009.
- [9] F. J. Massey, "The Kolmogorov-Smirnov Test for Goodness of Fit," Journal of the American Statistical Association, vol. 46, no. 253, pp. 68-78, 1951, doi: 10.2307/2280095.