Tie Formation and Collaboration Patterns among Music Artists

Rafael Godinho

1 Introduction

Since the mid-2010s, streaming services have revolutionized the music industry, transforming the way music is consumed and distributed. With the rise of platforms such as Spotify, Apple Music, and YouTube Music, streaming has emerged as the primary source of music consumption within the global recorded music market. This paradigm shift has had profound implications for tie formation and collaboration among music artists. As streaming services continue to dominate the music landscape, artists have recognized the potential for new exposure through collaborations. The ability to seamlessly access music worldwide has incentivized artists to seek out strategic collaborations that not only enhance their artistic expression, but also maximize their visibility and commercial success. Understanding the intricacies of these collaborations sheds light on the social dynamics within the industry while providing valuable insights into the creation and dissemination of music. The goal of this analysis is to delve into the exploration of tie formation and collaboration patterns among music artists. By analyzing network data and employing statistical modeling techniques, I seek to uncover the underlying factors that drive artists to form ties and collaborate with one another.

2 Data

2.1 Data Collection

The complete dataset was constructed using data sourced from Spotify, a prominent music streaming platform. This data was sourced from Spotify in 2020. It encompasses a collection of weekly charts featuring the top 200 most streamed songs in various countries and territories where Spotify is available. The data was collected through Zenodo.com. The source of the data was collected for a conference paper at ISMIR 2020. It was collected by Gabriel P. Oliveira, Mariana O. Silva, Danilo B. Seufitelli, Anisio Lacerda, and Mirella M. Moro.

2.2 Dataset Description

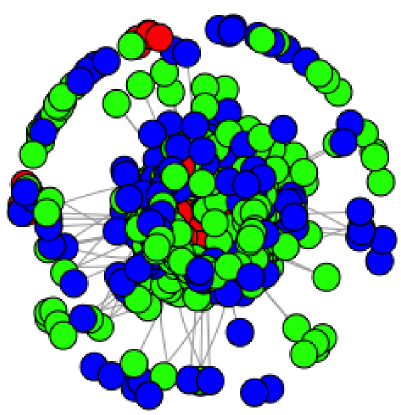
There are two subsets of the dataset that were utilized for this analysis. The first subset is an Artist Collaboration Network. The network of collaborating artists was derived from the Spotify Charts for the U.S market from 2017 to 2019. A collaboration between artists is established when they jointly work on a hit song, indicated by its inclusion in the Spotify Top 200 Chart. The subset is comprised of four main variables: "Artist1" representing the primary artist's name, "Artist2" representing the collaborating artist's name, "Count" indicating the number of hit songs in which both artists collaborated, and "Song_ids" providing the Spotify ID for each hit song involving both artists.

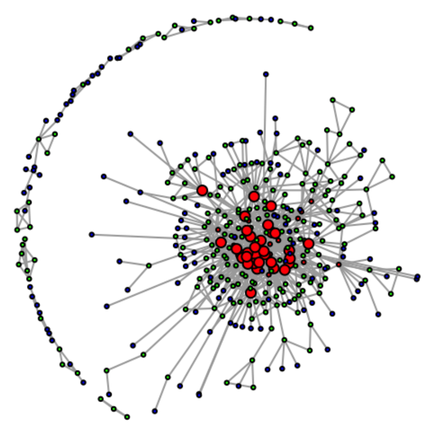
The second subset, known as the Artist Dataset, contains additional information about the artists who created the hit songs on the US Top 200 Spotify Charts from 2017 to 2019. The Artist Dataset includes three key variables: "Name" representing the artist's name, "Popularity" indicating the artist's overall popularity on a scale from 1 to 100, and "Genres" encompassing a list of the artist's music genres.

The data population represented in this analysis consists of collaborating artists who contributed to hit songs within the US market. This dataset holds several strengths, including its weekly charts of the 200 most streamed songs over multiple years as well as a comprehensive artist collaboration network based on their success. These strengths provide valuable insights into the dynamics of the music industry and artistic collaborations. However, it is important to acknowledge a limitation of the data, which stems from its exclusive reliance on Spotify. Consequently, the dataset may not fully capture the entirety of the music consumption landscape, as it does not account for other streaming platforms, offline music consumption patterns and countries where Spotify is not currently available.

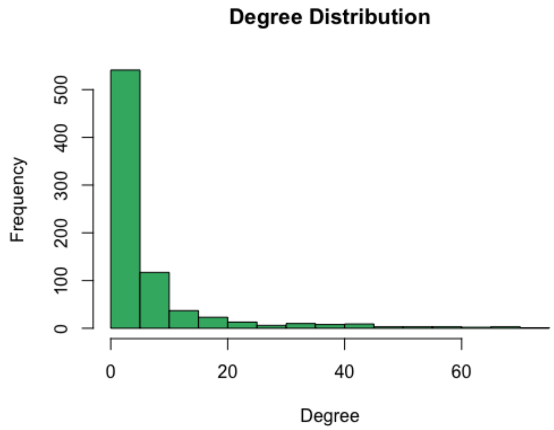
3 Network Visualizations

The Artist Collaboration Network analyzed in this study comprises 2666 unique collaborations involving 779 different artists over a three-year period. The network is represented as an undirected graph, and the nodes are color-coded based on their degree, a measure of the number of collaborations an artist has. Blue nodes represent artists with a degree of 1, indicating





collaborations with a single artist. Green nodes represent artists with a degree between 2 and 8, while red nodes represent artists with a degree of 9 and above. These degree categories were

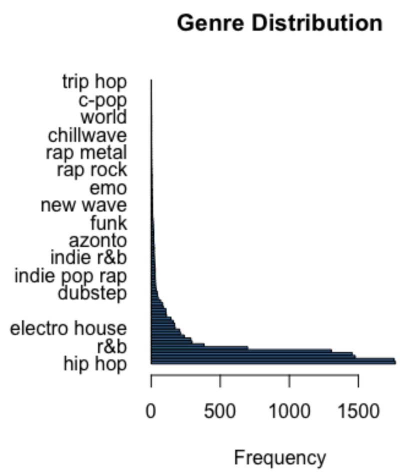


chosen considering that many artists have only had one song that made it on the charts, making the network representative of this pattern.

The histogram of Artist Degree Distributions reveals that a vast majority of artists have a degree less than 9, indicating a limited number of different collaborations. Artists with degrees exceeding 9 are in a significantly higher percentile compared to the rest of the network, signifying extensive collaboration with different artists. Notably, the artists with the highest degrees include Quavo at 71 collaborations (pop rap, rap, trap, hip hop), Nicki Minaj at 70 (pop, pop rap, dance pop, hip hop), Future at 69 (pop rap, rap, trap, hip hop), Travis Scott at 69 (rap) and Gucci Mane at 65 (pop rap, rap, trap, hip hop). Interestingly, these highly collaborative artists differ from the most popular artists in the network, which consists of Bad Bunny at a rating of 100 (latin, reggaeton, trap), The Weeknd at 98 (r&b, pop), Drake at 98 (rap, pop rap, pop, hip hop), J Balvin at 97 (latin, reggaeton) and Billie Eilish at 95 (pop, electropop). This divergence can be attributed to differences in music consumption patterns between the US market and the rest of the world.

The network was visualized using the Kamada-Kawai layout, revealing a prominent cluster in the central part of the network, predominantly occupied by high-degree nodes. Surrounding this central cluster are artists with a moderate number of collaborations (2 to 8). The outer ring of the network comprises artists who are not directly connected to the main subgroup. This outer ring likely consists of smaller artists who are yet to establish themselves in mainstream music. Examining the most successful artist collaborations, Lil' Baby and Gunna (rap, trap, hip hop) top the list with 34 hit songs, followed by Quavo and Travis Scott with 25 hit songs, and 21 Savage (rap, trap, hip hop) and Metro Boomin (pop rap, rap, trap, hip hop) with 21 hit songs. These collaborations demonstrate the prevalence of rap, trap, hip hop and pop rap genres.

The Genre Distribution barplot illustrates the distribution of genres among the collaborating artists. The most popular genres among the collaborations are hip hop, rap, pop rap, trap, and pop. The barplot exhibits an exponential-like distribution, with the top genres appearing



in over half of all collaborations, while the majority of genres are significantly lower.

The structure of data in these formats resembles a scale-free network as described by Barabasi & Bonabeau. A scale-free network can be characterized by a power-law distribution of node degrees, meaning that a few nodes have a large number of connections while most nodes have only a few connections. In this context, a scale-free network suggests that there are a few highly influential artists who have numerous collaborations, while the majority of artists have significantly fewer collaborations. This reflects the hierarchical nature of the music industry, where a small number of prominent artists dominate the collaborations and have a significant impact on the overall network structure.

4 Analysis

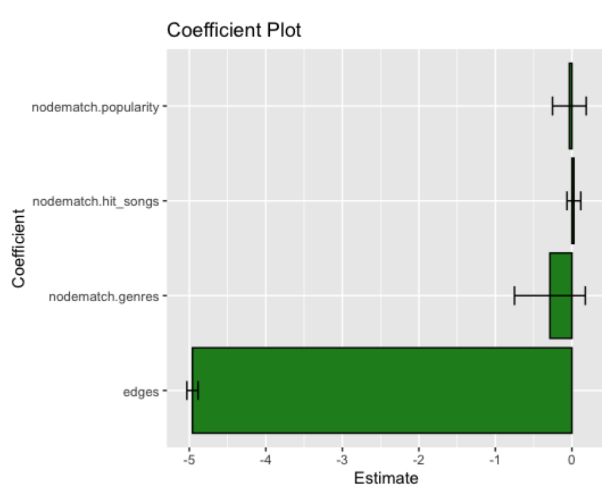
4.1 Method

To understand the underlying mechanisms of tie formation in the artist collaboration network, an exponential random graph model (ERGM) was employed for the analysis. The ERGM is a statistical model that allows us to examine the probability of ties between artists based on various attributes. In our case, we included the edges (density), number of hit songs, artist popularity and genre similarity as the predictors of tie formation. The ERGM model was chosen for this analysis because it provides a flexible framework for modeling tie formation in networks. It allows us to capture the complex interplay between different factors and assess their impact on collaboration patterns. By fitting the ERGM to the artist collaboration network, I can estimate the effects of the included attributes and gain insights into mechanisms that drive tie formation.

To prepare the data for analysis, I created a network object using the artist collaboration data. I also merged the artist information dataset with the collaboration data, incorporating the vector attributes as well. This allowed me to analyze the influence of these attributes on tie formation and collaboration patterns. Nodematch was utilized in the ERGM model to better examine the effects of attribute similarity on tie formation. This is done by comparing the attributes of each pair of nodes in the network. The term calculates a match statistic that quantifies the degree of similarity or dissimilarity between the attribute values of the nodes.

4.2 Results

The estimated coefficient for the edges term is -4.95892. This negative coefficient suggests that there is a strong tendency for artists to collaborate less frequently than expected by chance alone. The coefficient is also highly significant at $p < 0.001$, which indicates a robust finding. The estimated coefficient for the hit songs term is 0.02579. This coefficient suggests that artists who have collaborated on hit songs in the past are slightly more likely to form new collaborations. However, the coefficient is not statistically significant at $p = 0.568$, indicating that the effect may not be substantial. The estimated coefficient for the popularity term is -0.03315. This coefficient suggests that there is a slightly negative relationship between artist popularity and tie formation. However, the coefficient is also not statistically significant at $p = 0.763$, indicating that artist popularity may not strongly influence collaboration patterns. The estimated coefficient for the genre term is -0.28756. This coefficient suggests that artists who share similar genres are less likely to collaborate. However, the coefficient is also not statistically



significant at $p = 0.214$, indicating that genre similarity may not be a major driver of tie formation. The coefficient plot indicates a strong statistical significance for the edges term, while the genres term, although not statistically significant, remains noticeably higher than the hit songs and popularity terms.

The AIC value of 25450 and the BIC value of 25492 indicate that the model has relatively good fit when considering the complexity of the model. The goodness-of-fit statistics indicate that the model provides a significantly better fit to the data than the null model. The null deviance is much higher than the residual deviance, indicating that the included predictors improved the model's fit.

5 Conclusion

In the case of the edges term, the negative sign of the coefficient suggests that as the number of edges in the network increases, the likelihood of forming new ties between music artists decreases. This implies that there is a tendency for tie formations to be less frequent as the overall connectivity of the network increases. In the context of the music industry, this result may suggest that as the number of collaborations and connections between artists increases, the rate at which new collaborations are formed decreases. This could be due to the diminishing novelty of new collaborations in an already established and well-connected network. It also may indicate that artists are more selective or strategic in choosing their collaborators as the network becomes more saturated. Artists likely maintain the established relationships they already have, rather than actively seek out new ones. This may be due to shared artistic vision, chemistry, trust or previous successful collaborations

As for the rest of the terms, the analysis revealed that there is no significant association between the number of hit songs, shared genres, or popularity of two artists and the likelihood of forming a tie/collaboration. These results may imply that tie formations in the music industry are driven by other factors such as personal connections, shared interests or other social and professional networks. It is also important to consider other factors such as personal relationships, artistic preferences, management decisions and industry dynamics that may influence tie formations within the music industry.

6 Remarks

There are a multitude of ways that I'd like to expand upon this project with more analysis or better data. Firstly, I'd like to utilize the global Spotify charts alongside the charts for other respective countries such as the UK or Brazil to see if there are any discernible differences in regards to tie formation. This analysis would also integrate data from other music streaming platforms such as Apple Music and YouTube Music to provide a more comprehensive analysis of the attributes alongside the inclusion of new ones such as song release dates. The inclusion of more data would hopefully reduce the variability of my analysis. I'd also like to consider studying the dynamics of the network over time. Instead of merging data from multiple years, I'd analyze each year separately and examine how the network structure and attribute correlations change over time. Finally, I'd like to develop predictive models to forecast future collaborations or attribute values based on historical patterns. This could be valuable for identifying emerging trends, predicting successful collaborations or to simply form a better understanding of the factors that contribute to artist popularity and their hit songs.

The findings of this analysis have the potential to contribute to various areas within the music industry and beyond. Firstly, understanding tie formation and collaboration patterns can aid music industry professionals, such as record labels, managers, and booking agents, in making informed decisions regarding artist collaborations, tour planning, and resource allocation. Additionally, the findings can offer valuable insights into the dynamics of cultural exchange, artistic innovation, and the diffusion of musical trends within and across genres.