Analyses of Musical Success based on Time, Genre and Collaboration

Gabriel P. Oliveira, Anisio Lacerda, Mirella M. Moro

[1]Programa de Pós-Graduação em Ciência da Computação (PPGCC)
Universidade Federal de Minas Gerais (UFMG) – Belo Horizonte, Brazil

{gabrielpoliveira,anisio,mirella}@dcc.ufmg.br

Abstract. *Music holds a significant position in global culture, as it is one of the world's most important and dynamic cultural forms. With the vast amount of music-related data available on the Web, new opportunities emerge for extracting knowledge and benefiting different music segments. In this work, we perform a data-driven analysis to investigate musical success from a genre-oriented perspective. Specifically, we model both artist and genre success timelines to detect and predict continuous periods with higher impact. We also build success-based genre collaboration networks to detect collaboration profiles directly related to success. Furthermore, we use data mining techniques to uncover exceptional genre patterns in the networks where the success deviates from the average. Our findings show that studying genre collaboration is a powerful way to assess musical success by describing similar behaviors within collaborative songs. Overall, our work contributes to both the academy and the music industry, as we shed light on the underlying factors of the science behind musical success.*

1. Introduction

Music is not just one of the world's most important cultural industries, but also one of the most dynamic. With the sheer volume of music content available on the Web, new challenges arise for the music industry every day. The massive volume of complex data generated by music digital platforms such as Billboard and Spotify presents numerous research opportunities across various domains. For instance, leveraging data about songs, their characteristics, and the social interactions about them reveal meaningful insights into music consumption, user behavior, social dynamics, emotional responses, and industry-related aspects. Given the increasing volume of data and the complexity of such tasks, Music Information Retrieval (MIR) emerges as an interdisciplinary research area combining Computer Science and Musicology to extract meaningful information from musical content. Research in MIR uses knowledge from Machine Learning, Data Mining, and Data Science in many applications, from assisting in musical composition [Suh et al. 2021] to predicting potential success [Cosimato et al. 2019].

The dynamic nature of the music industry can directly influence the behavior of artists' careers. That is, an artist's career can suffer ups and downs depending on the current market moment. At a higher level of abstraction, the same fluctuating behavior happens for musical genres. From the 1960s to the 1980s, *soul* and *rock* genres dominated the music scene, with Stevie Wonder, Aretha Franklin, The Beatles, and Queen being some of the greatest artists of this period. A substantial change in musical genre preferences marked the 1990s, mainly due to technological advances, such as Internet

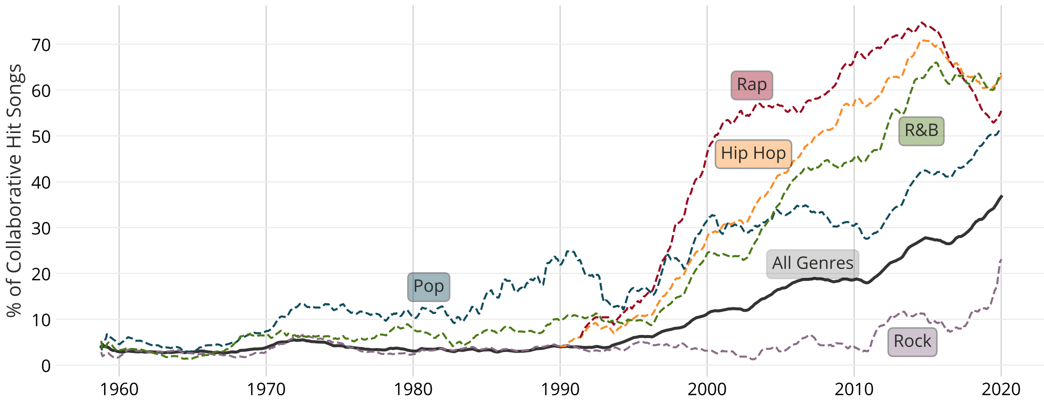

Figure 1. Historical frequency of collaborative hit songs for selected genres on Billboard Hot 100 Chart (1958 - 2020).

popularization. From then on, *pop* and *rap* conquered space on the charts and became protagonists at the beginning of the 2000s. Britney Spears, Eminem, Beyoncé, and Drake are examples of artists of such genres.

Indeed, as the music industry becomes more complex and competitive, artists are encouraged to reinvent strategies to maintain their presence in the market and reach new audiences. Thus, artist collaboration has grown into one of the main tactics to promote new songs. Such connections usually help artists bridge the gap between styles and genres, overlapping new fan bases and consequently increasing their numbers. In such a way, several studies approach the factors behind musical success, creating an emerging field within MIR called Hit Song Science (HSS). Collaboration-aware studies then become promising, as successful artists are more likely to have a high degree of collaboration in success-based networks [Silva et al. 2019]. In fact, there is strong evidence that factors leading to an ideal musical partnership can be understood by exploring collaboration patterns that directly impact its success [Bryan and Wang 2011].

The genre perspective is very important when analyzing the impact of collaborations in musical success, as each genre has a distinct audience that behaves in its own way. Figure 1 shows this phenomenon and highlights the growing trend in the number of collaborations within Billboard Hot 100 Charts. Although the general curve increases over time, genres such as *pop* and *R&B* are more collaborative than others (e.g., *rock*). This contrast can be explained by the intrinsic nature of each music genre, since *pop* and *R&B* artists frequently collaborate with the *rap* community, mainly as featured artists. Also, partnerships involving *pop* music may take place not only through intra-genre collaborations but also through inter-genres, bringing an additional dimension to their songs. Hence, as this creative industry changes, it becomes more unpredictable; and doing both predictive and diagnostic analyses in such a context remains challenging.

Devil's Advocate Perspectives. "What makes a song successful" is a billion-dollar question that has moved the music industry. The answer may involve a complex mix of many perspectives from psychological ones (emotions, musical taste, personal history, etc.) to economic and strategic (e.g., high exposure of its singer/band on media), social (e.g., everyone is listening to it) and intrinsic music features (beat, tone, lyrics, etc.). Although

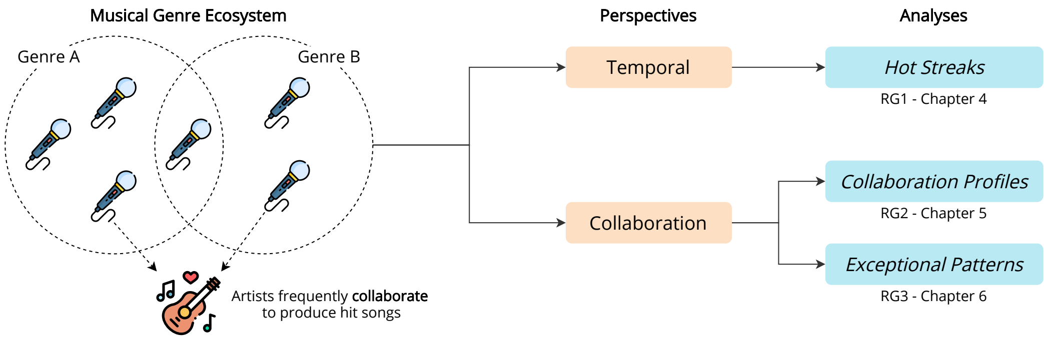

Figure 2. Analyses conducted in this work, according to the Research Goals.

true, such arguments have not stopped companies and startups from creating and selling their solutions (e.g., Polyphonic HMI and MixCloud), nor have they stopped the ever-growing MIR research field exploring solutions for almost two decades (see Section 3).

2. Research Goals and Contributions

Remaining an industry of creative growth, it is only natural for music (i.e., all musical scene members) adapting to new conditions and redefining its layout. Not surprisingly, the Grammy[1] categories were tightened (from 109 to 78, in 2012) as a result of music's dynamic nature. That is, the notion of categories and genres are blurred as never before. In addition, as the collaboration phenomenon becomes stronger over the years, it is necessary to explore all factors that make it so relevant nowadays. Therefore, this work aims to **analyze artist collaboration under a genre perspective to better understand how the genre connections impact musical success.** We do so by exploring the musical genre ecosystem in temporal and collaboration perspectives (Figure 2). Therefore, we assess such an objective through three distinct Research Goals (RGs):

RG1. Understand the temporal evolution of both artist and genre careers, by identifying and predicting periods of high impact in such careers, i.e., hot streaks (Section 4);

RG2. Analyze the dynamics of cross-genre connections by detecting collaboration profiles in success-based networks, i.e., connections formed by genres of artists who cooperate and create hit songs (Section 5);

RG3. Mine frequent genre patterns within hit songs in recent years, i.e., investigating the relationship between combining different genres and musical success (Section 6).

3. Related Work

Hit Song Science (HSS) tackles the problem of predicting the popularity of a given song, and it is also an emerging field within MIR. Thus, different studies analyze the impact of acoustic and social features in musical success. In the early years of HSS, only acoustic features (i.e., the internal technical aspects of a song, such as timbre, mode and key) were assessed by researchers [Dhanaraj and Logan 2005]. Nonetheless, as the Web became popular and widely adopted, social interactions were included as features in prediction models [Cosimato et al. 2019]. Understanding musical aspects can

[1]Grammy Awards: https://en.wikipedia.org/wiki/Grammy_Award

also be genre-dependent, and this also reflects in the musical success. Therefore, several studies in HSS use genre information in their models to understand the life trajectory of songs [Shin and Park 2018] and summarize the acoustic content of a song [Ren and Kauffman 2017]. Overall, there is strong evidence that music genre may influence musical success, and such information lead to improving the performance of success prediction models [Abel et al. 2010, Askin and Mauskapf 2017].

Moreover, Silva et al. [2019] address collaboration as a key factor in success, using topological properties to detect relevant profiles in artist networks. In a later study, the causality between collaboration and success is addressed [Silva and Moro 2019], increasing the knowledge and reinforcing the relevance of the collaboration phenomenon in the musical scenario. In fact, such an approach is novel and promising in HSS, but it is restricted to the artist and song levels. In addition, these and most of the aforementioned studies regarding musical success only consider data from American charts, mainly Billboard Hot 100. This may be due to the ease of obtaining data but it may not reflect the whole global scenario, as each country has its own distinct behavior when consuming music, which includes preferred artists and genres.

Therefore, one of the contributions of this work is to introduce a model to describe artists' and genres' success timelines by identifying and characterizing periods with success above the average (i.e., hot streaks). In addition, studying collaboration from a genre perspective may reveal important information on how artists from different communities team up to make a new hit song. To the best of our knowledge, we are the first to build a success-based genre network, investigating collaboration profiles over time and mining exceptional patterns within it, going deeper into the potential intrinsic factors that make up a successful collaboration. Likewise, our approach considering several regional markets makes this work more realistic, as local engagement shapes the global environment. We combine a precise heterogeneous data collection with proper modeling to enhance further data analysis by scientists and record labels CEOs.

4. Hot Streaks in Musical Careers (RG1)

Careers tend to have phases of high productivity, reaching peaks. Hot streaks (HS) is the term commonly used for continuous periods of success above normal. Overall, the music industry is as dynamic as it is a crucial part of the entertainment world. Within so much uncertainty, a clear fact is: when an artist is at a hot streak, such an artist is also at the most profitable moment of a career. In this section, we identify and predict hot streaks in the music scene, defined by high-impact bursts occurring in sequence within artist careers.

Data Collection. We collect all Billboard Hot 100 charts from August 11, 1958 to August 22, 2020 (data collection time). We enrich our dataset by collecting data from Spotify. Specifically, we obtain artist genres and debut date, as well as acoustic features for each song, such as key, mode, and energy. Our final dataset, called MUHSIC (Music-oriented Hot Streak Information Collection) is composed of 3,238 weekly charts containing 24,540 distinct songs from 6,248 artists belonging to 998 music genres [Oliveira et al. 2021a].

Time Series Modeling. Success in the music industry has a temporal structure, as the audience tastes change over time. For each artist, we build their time series from the debut date (i.e., date of the first release obtained from Spotify) to the last chart collected. Thus, each point in the time series represents the success of such an artist in a given

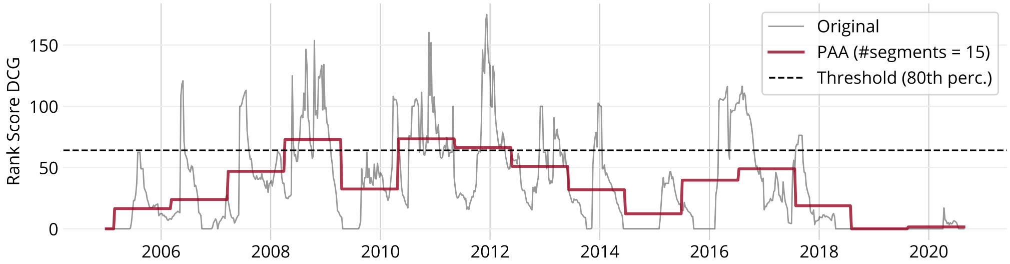

Figure 3. Piecewise Aggregate Approximation (PAA) applied to Rihanna's success time series (2005–2020). Periods above the threshold are hot streaks.

Table 1. Classification evaluation results. Values are presented with a 95% confidence interval (CI) and bold values indicate that a classifier is statistically better.

	LR	LinearSVC	Perceptron	SGD
Accuracy	**0.90993 ± 0.00001**	0.90791 ± 0.00005	0.88236 ± 0.01954	0.90505 ± 0.00190
Precision	0.73536 ± 0.00004	0.72764 ± 0.00015	0.67466 ± 0.07060	0.73035 ± 0.00508
Recall	0.78935 ± 0.00004	0.78977 ± 0.00030	0.77301 ± 0.09725	0.75886 ± 0.01923
F1 Score	**0.76140 ± 0.00002**	0.75743 ± 0.00015	0.70383 ± 0.03769	0.74405 ± 0.00828
F1 Weighted	**0.91116 ± 0.00001**	0.90935 ± 0.00005	0.88551 ± 0.01623	0.90572 ± 0.00229

week, according to the Hot 100 chart. Analyzing musical genres' success over time is also one of the goals of this work, and we can build genre success time series based on data obtained from Spotify. First, we assign artists' genres to their songs, as the songs themselves do not have such information. Then, for each week, we aggregate all songs from artists belonging to a given genre that appear on that week's chart.

Hot Streak Detection. After modeling success in musical careers, we detect periods of higher impact (i.e., hot streaks) in them, following recent research on this subject [Garimella and West 2019]. First, we use Piecewise Aggregate Approximation [Keogh and Pazzani 2000] to delimiter periods within careers. Then, we define a hot streak as the periods in which the success is above a certain threshold, obtained from the artist/genre career itself. Figure 3 illustrates Rihanna's career and her hot streaks, the first one from April 2008 to April 2009 (*Disturbia* and *Take a Bow*) and the second from May 2010 to May 2012 (*What's My Name*, *Only Girl (In the World)* and *We Found Love*).

Hot Streak Prediction. We model the hot streak prediction as a *binary classification* task in which, for a given week, an algorithm predicts whether it belongs to a hot streak period in a time series or not. Here, we consider genre time series for the prediction because previous research provide evidence that a luck component plays an important role within success in individual careers [Janosov et al. 2020, Sinatra et al. 2016]. On the other hand, genre careers are more stable and have well-established hot streak periods, providing examples of both hot streak and non-hot streak periods for the learning algorithms.

Figure 4 illustrates our classification model, called *MHSBC – Music-oriented Hot Streak Binary Classification*. For each week in the genres' time series, MHSBC calculates a set of features (i.e., genre-related, artist-related, and song-related) describing all songs from one genre which are in the week's chart. The genre information becomes a

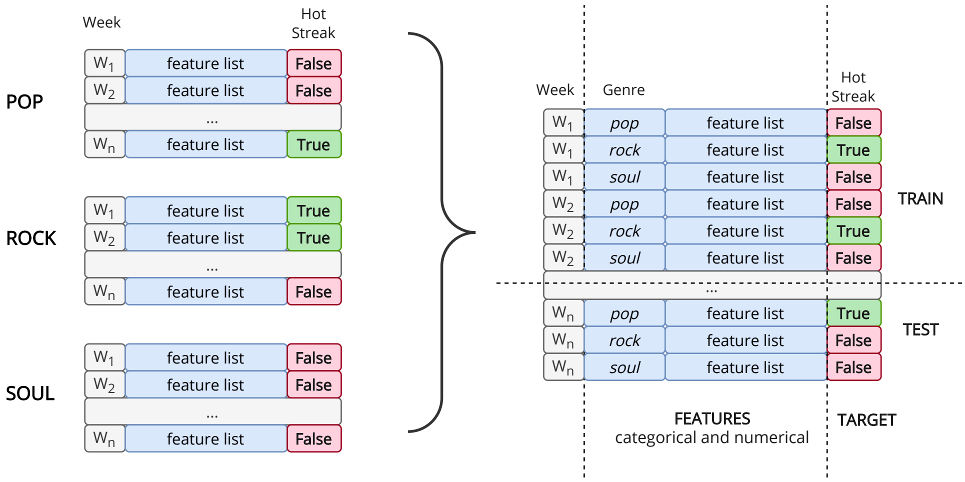

Figure 4. Music-oriented Hot Streak Binary Classification from genre time series.

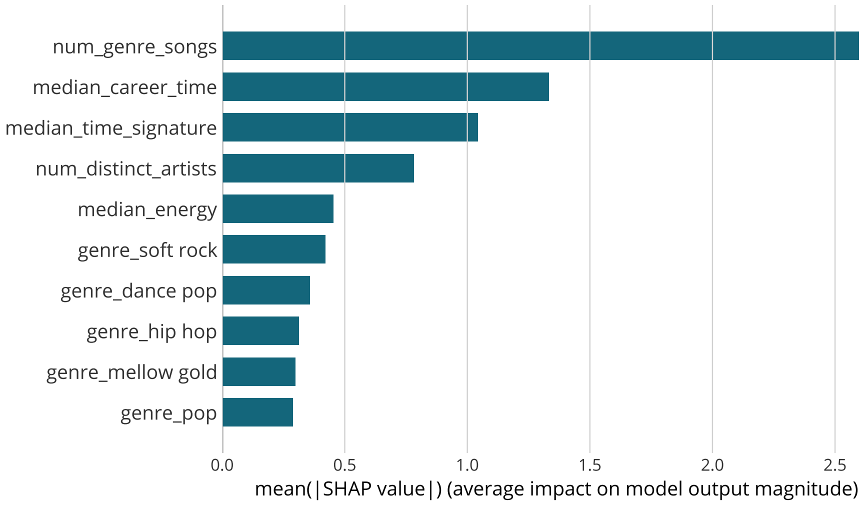

Figure 5. Features with the highest absolute mean SHAP values.

categorical feature in this final set. Hence, in MHSBC model, each instance describes the set of songs from a given genre which entered the Hot 100 chart in that week.

Table 1 presents the results for all classifiers considered in this study. All classifiers outperform the baseline, which predicts the most frequent class ($accuracy = 0.85$). Thus, our model is better than simply guessing, as all classifiers have higher accuracy values. The baseline does not provide F1-score, as it does not make predictions (i.e., it simply returns the majority class for every instance). Considering the classifiers, Linear Regression (LR) provides the best results, with average accuracy of 0.91 and average F1-score of 0.761, which are significantly higher (95% confidence interval) than the other algorithms. Hence, we choose it as the one for MHSBC.

Feature Importance. Using machine learning methods for hot streak prediction produces results with high accuracy, but understanding why and how a model makes a certain prediction can be as crucial as the outcome itself. Here, we use SHAP (SHapley

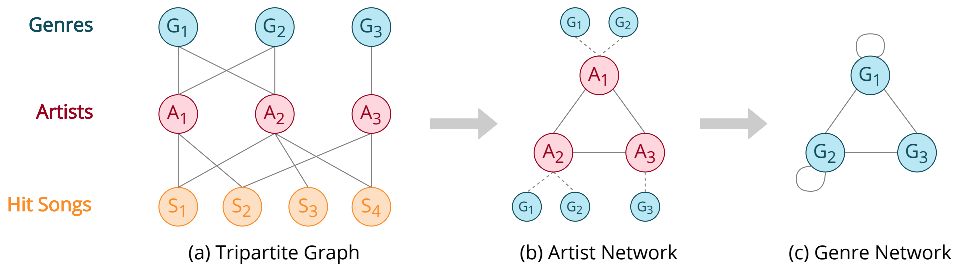

Figure 6. Reduction from the tripartite (a) to the one-mode Genre Network (c). The intermediate step is an Artist Network with genre information (b).

Additive exPlanations) [Lundberg and Lee 2017] in our classification model to allow its interpretability. In short, SHAP is a game-theoretic approach to explain the output of a machine learning model, assigning for each feature an importance value for a particular prediction. Figure 5 is a summary plot with the top 10 features with the highest impact on hot streak prediction. The result shows features such as the number of songs, median time signature, median career time, and the number of distinct artists are the most significant to our model, thus having high predictive power. The descriptive genre features (e.g., *genre_pop* and *genre_rap*) also appear in the ranking, but their average SHAP values are close to zero, requiring further investigation.

Overall, understanding the dynamics around hot streak periods and also to predict their occurrence is relevant not only to the scientific community but to the music industry as a whole. For the first, it may contribute to developing more complex models, while for the latter it helps to describe the listeners' behavior and success trends over artists and genres. Therefore, both musicians and record labels may orientate their future releases to achieve or to maintain their success levels. In short, the real value of identifying hot streaks is in revealing the fundamental patterns that govern individual careers.

5. Collaboration Profiles in Genre Networks (RG2)

So far, we have assessed musical careers by analyzing hot streaks of artists and bands, corresponding to phases in which the individual success is above-normal. Genre appeared as a relevant feature, as there are different patterns of hot streaks according to the genre. Now, we add a dimension to our analyses by considering the collaborations that connect different genres. Hence, our hypothesis is that success is not only related to the performance on charts, but also to the genre connections that make hit songs.

Data Collection. We collect global and regional charts from 2017 to 2019, considering eight of the top 10 music markets according to IFPI[2]: United States (1st), Japan (2nd), United Kingdom (3rd), Germany (4th), France (5th), Canada (8th), Australia (9th), and Brazil (10th). We also use Spotify API[3] to gather information about the hit songs and artists present in the charts, such as all collaborating artists within a song and their respective genres, which is the core of this work. Our final dataset, called MGD (Music Genre Dataset) contains 1,370 charts from 156 weeks, comprising 13,880 hit songs and 3,612 artists from 896 different genres [Oliveira et al. 2020c].

[2] Spotify was not available in South Korea (6th) and China (7th) as of May 2020 (collection date).
[3] Spotify API: https://developer.spotify.com/

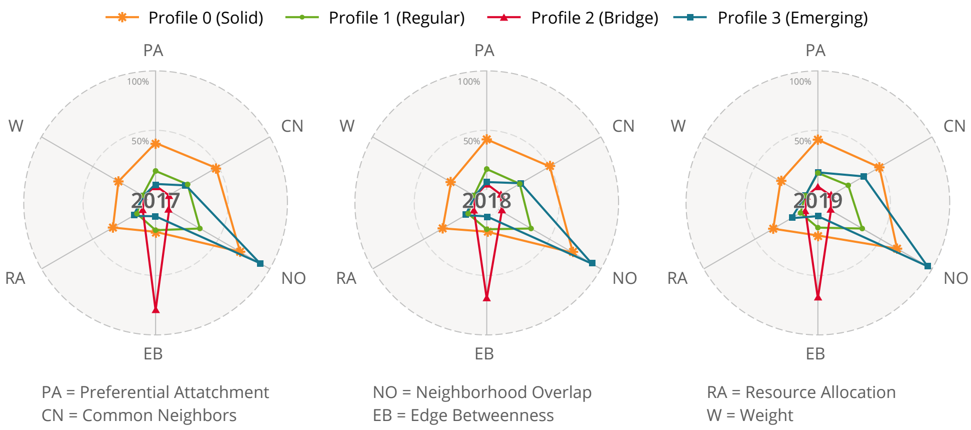

Figure 7. Collaboration profiles for all markets (2017-2019).

Genre Collaboration Network. A Collaboration Network is usually modeled as a graph formed by nodes (vertices) that may be connected through edges. Here, we model music collaboration as a tripartite graph (Figure 6a), in which nodes are genres, artists, and hit songs. Collaborative hit songs are sung by two or more artists, regardless of their participation (e.g., a *feat.* or a duet). We then reduce the tripartite model into a one-mode network in which nodes are exclusively genres. However, such a reduction is only possible by an intermediate step: building the artist collaboration network (Figure 6b). In such a network, two artists are connected when both collaborate in one or more hit songs. The genres are not lost, as they are linked directly to the artists. We may now build the final network by connecting the genres of artists who collaborate in the artist network. The edges are undirected and weighted by the number of hit songs involving artists from both genres (Figure 6c). Also, self-loop edges are allowed, as there are hit songs from artists of the same genre. With nine markets (global and eight countries) during three years (from 2017 to 2019), we analyze 27 networks.[4]

Collaboration Profiles. We now present our approach to uncover significant factors that compose a successful music genre collaboration. Inspired by [Silva et al. 2019], we first extract information from the success-based networks by evaluating six edge-dependent metrics. We perform an Exploratory Factor Analysis on such metrics to define factors, and then perform a cluster analysis using the DBSCAN algorithm. Overall, four distinct clusters were detected in our networks. Figure 7 shows the radar charts with the mean values of each market present in that profile. To summarize the characteristics of the collaboration profiles, we name each as follows.

- ***Solid*** *Collaboration*, composed of well-established collaborations between most popular genres (super-genres), which have been going on for decades;
- ***Regular*** *Collaboration*, composed of the most common collaborations in all markets, which are very similar to solid collaborations but not as engaged;
- ***Bridge*** *Collaboration*, composed of collaborations with high influence, representing bridge-like connectors between two areas of a network (mostly between divergent music styles). Such collaborations may be possible sources of investment to increase connectivity and strengthen ties among different audiences; and

[4]All networks can be visualized in https://bit.ly/proj-Bade

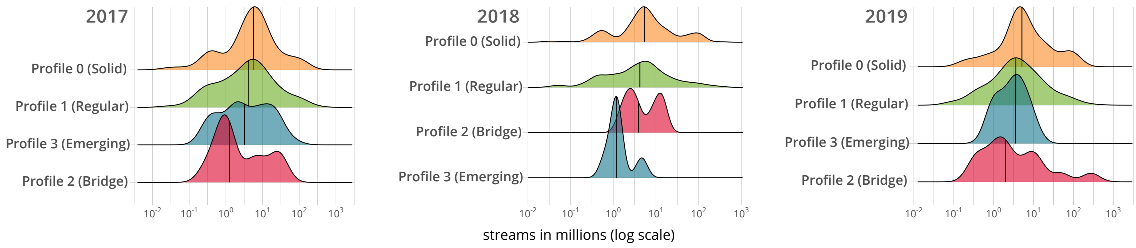

Figure 8. Density ridgeline plots of streams in millions for each cluster (log scale). Clusters are sorted by their median stream values (darker vertical lines).

Table 2. Total number of *intra-* and *inter*-genre collaborations in each profile.

Collab	*Solid* 2017	2018	2019	*Regular* 2017	2018	2019
Inter-genre	140 (49%)	125 (42%)	103 (51%)	1,828 (99%)	1,916 (98%)	2,165 (94%)
Intra-genre	145 (51%)	174 (58%)	99 (49%)	23 (1%)	34 (2%)	128 (6%)
Collab	***Bridge* 2017**	**2018**	**2019**	***Emerging* 2017**	**2018**	**2019**
Inter-genre	10 (100%)	7 (100%)	16 (100%)	3 (7%)	1 (17%)	0 (0%)
Intra-genre	0 (0%)	0 (0%)	0 (0%)	40 (93%)	5 (83%)	7 (100%)

- ***Emerging*** *Collaboration*, formed mainly of collaborations between regional genres. We propose the term *emerging* because such a profile can be seen as a transition phase for beginners, until they establish their fan bases.

Success Analysis. To better understand each profile, we analyze the distributions of success rate, and then the number of *intra-* and *inter*-genre collaborations for each profile. Here, we define the success rate as the average of streams of songs belonging to the genres that compose the collaboration (edge) in that year. Figure 8 shows the success density ridgeline plots for each profile, indicating that Profiles *Solid* and *Regular* are composed of the most successful music genre collaborations, on average. With results from Table 2, in general, the most successful profiles are those composed of more inter-genre collaborations. Such a result may indicate a strong correlation between musical success and inter-genre collaborations. Indeed, by teaming up with one (or more) person of a different musical style in a song, both artists may draw from one another's fan bases; i.e., they may promote themselves to new public who could increase their fan base and audiences.

Overall, detecting genre collaboration profiles is a powerful way to assess musical success by describing similar behaviors within collaborative songs from multiple angles. Our findings may act as base material for further research tasks, e.g., prediction and recommendation. The former enables predicting the success of a given song/artist/album, while the latter can be used to point out potentially successful genre/artist collaborations. In fact, music industry CEOs may maximize expected success by properly investing in potential artist/genre collaborations. Finally, artists may also profit by identifying the most suitable partnerships to lead the album to early stardom.

Table 3. Frequent patterns in global and non-English speaking markets (2019).

Market	Pattern	Support	Market	Pattern	Support
Global	('dance pop', 'pop')	0.271	*Brazil*	('brazilian funk', 'pop')	0.177
	('latin', 'reggaeton')	0.173		('electro', 'brazilian funk')	0.102
	('hip hop', 'trap')	0.172		('sertanejo', 'brazilian funk')	0.097
	('rap', 'hip hop')	0.168		('electro', 'pop')	0.080
	('rap', 'trap')	0.151		('trap', 'hip hop')	0.064
France	('hip hop', 'pop')	0.584	*Japan*	('j-rock', 'j-pop')	0.283
	('rap', 'hip hop')	0.449		('other', 'j-pop')	0.140
	('rap', 'pop')	0.423		('anime', 'j-pop')	0.138
	('rap', 'hip hop', 'pop')	0.393		('dance pop', 'pop')	0.133
	('francoton', 'pop')	0.174		('r&b', 'j-pop')	0.108

6. Exceptional Genre Patterns in Hit Songs (RG3)

Given the diversity of collaborations between artists from several genres, it becomes challenging to conduct predictive and descriptive analyses in such a context. For example, it may be relevant to record labels to uncover frequent genre collaborations with higher success to plan future song releases. In this section, we go further in the study of genre collaborations by using our networks and collaboration profiles to mine exceptional patterns of musical genres in hit songs, i.e., to verify if there is a relationship between combining different musical genres and success. We use two data mining techniques in our experiments: Frequent Itemset Mining and Subgroup Discovery [Zaki and Meira Jr. 2014].

Genre Frequent Patterns. We focus on finding the most frequent genre associations by applying the Apriori algorithm from the set of hit songs in each musical market from the Music Genre Dataset (MGD). We define the transactions of our task as hit songs, whose items are the musical genres of each artist who sing them. As language is crucial for listening to music, we divide our eight regional markets into two distinct groups: English and non-English speaking countries. We then perform our analyses comparing the countries with each other and the patterns found in the global charts, which is an aggregation of all territories in which Spotify is available.

For instance, the analysis of frequent genre patterns for non-English speaking countries reveals a strong regional component in most countries. Table 3 presents the five most frequent genre associations in 2019 for three countries: Brazil, France, and Japan. All such countries have patterns with regional rhythms, such as *francoton* and *brazilian funk*. Still, Japan stands out, as all five patterns have regional styles, with main genres including *j-pop*, *j-rock* and *anime*. Besides, our results reveal the absence of genres such as *hip hop* and *rap* in Japan, which are present in all other markets. In all countries, the presence of local genres increased over time, revealing a tendency of the population to value their own culture and consequently promote it globally.

Exceptional Genre Collaborations. Here, we perform a subgroup discovery (SD) [Klösgen and Zytkow 2002] in our networks to find exceptional genre collaboration patterns from hit songs, i.e., collaborations in which the success is above the average in the whole dataset. In our approach (summarized in Figure 9), for each market and year, we consider the network edges as the instances of the SD model, representing the collaboration between musical genres.The attribute set for each instance is composed of features

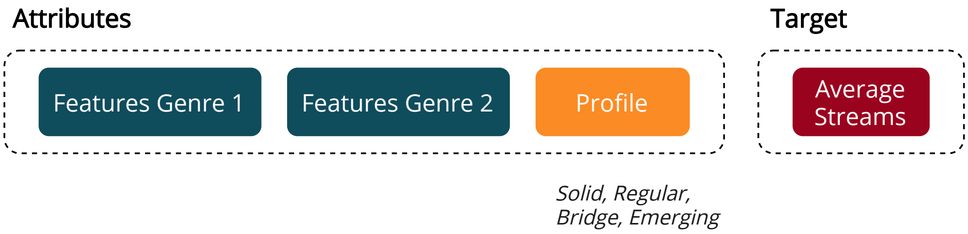

Figure 9. Representing the edges of Genre Collaboration Networks as instances of the Subgroup Discovery (SD) problem.

from the two genres of the respective edge (acoustic features obtained from Spotify), as well as the collaboration profile (*Solid*, *Regular*, *Bridge* and *Emerging*) for such an edge.

In short, all markets have subgroups with attributes and target distributions different from each other. However, when analyzing each subgroup individually, there are connections that repeat in some countries. For instance the connection between *dubstep* and *pop* in 2018 is within exceptional subgroups in five markets: Australia, Canada, Germany, UK, and the US. Except for Germany, all countries are English-speaking and share several cultural aspects. Besides, the social media advance provides a global platform where users can share and promote their musical tastes.

Recommending Promising Genre Associations. We now uncover association rules, which inform how items (i.e., music genres) are associated with each other. Association rules are represented with expressions from the type $A \rightarrow B$, representing the occurrence of an itemset B (i.e., consequent) given that A (i.e., antecedent) also happens. We define promising rules according to their lift value, which measures their level of surprise. Values above 1 mean that the consequent is much more likely to happen than expected, given the occurrence of the antecedent. In contrast, values below 1 represent the opposite. As we aim to find the most promising genre associations, we look for rules with high lift values. Table 4 present the three most promising rules for selected markets in 2019. The results for the global market reveal the strong association of regional genres such as *latin*, *reggaeton*, and *tropical*. In Brazil, the genre *pagode baiano* appears on 42.5% of the songs containing *brazilian funk* and *pop*. Thus, combining such genres increases considerably the chances of a song to reach Brazilian charts, as did Anitta in her songs *Onda Diferente* (with Ludmilla) and *Combatchy* (with Luísa Sonza, Lexa and MC Rebecca).

Performing diagnostic analyses is crucial in music, as it allows the understanding of some relevant aspects behind success. Following the findings from previous sections, our results reinforce the importance of analyzing regional markets, as they behave differently compared to the global scenario or even to the United States (i.e., the biggest music market in the world). For example, in the past few years, the world has seen local genres such as *reggaeton* and *k-pop* becoming extremely popular worldwide. Therefore, our findings provide benefits to artists and record labels, as they serve as a first step in developing strategies to promote their work across the world.

7. Conclusions

In the past few years, music has been transformed by digital technologies and analytics. Such a transformation produces a huge volume of data on how people relate to music and

Table 4. Most promising association rules in selected markets by lift value (2019).

Market	Rule	Lift	Confidence
	('latin', 'reggaeton') → tropical	7.922	0.468
Global	('latin') → tropical	7.821	0.462
	('reggaeton') → tropical	7.722	0.456
	('tropical house') → house	7.655	0.342
Australia	('tropical house', 'pop') → house	7.173	0.321
	('tropical house', 'pop') → electro	7.111	0.670
	('hip hop') → trap	6.187	0.434
Brazil	('brazilian funk', 'pop') → pagode baiano	5.473	0.425
	('hip hop') → pop rap	5.235	0.303
	('r&b') → soul	7.485	0.226
Canada	('dance pop') → tropical house	3.214	0.243
	('dance pop', 'pop') → tropical house	3.160	0.239
	('rap', 'pop') → hip hop	1.301	0.900
France	('rap', 'pop') → francoton	1.263	0.325
	('hip hop', 'pop') → rap	1.259	0.796
	('dance pop') → tropical house	5.909	0.400
Germany	('dance pop') → electro	5.908	0.338
	('dance pop', 'pop') → tropical house	5.824	0.394
	('r&b') → j-rap	8.067	0.228
Japan	('dance pop') → electro	4.348	0.283
	('dance pop', 'pop') → electro	4.284	0.279
	('rock') → indie rock	8.370	0.364
UK	('rock') → indie	6.216	0.231
	('pop rap', 'hip hop') → trap	5.682	0.660
	('pop rap', 'pop', 'rap') → r&b	2.990	0.291
USA	('pop', 'rap') → r&b	2.888	0.281
	('hip hop', 'pop') → r&b	2.878	0.280

affects the whole music ecosystem, i.e., listeners, artists, record labels, and other agents in this industry. This work provides an extensive diagnosis and relevant insights on how genres relate to musical success. Our results showed that the occurrence of hot streak periods depends not only on the songs present in the charts but also on information obtained from the genres and artists, such as collaboration. Moreover, by analyzing genre networks, we uncovered distinct collaboration profiles, which are an important tool to assess musical success, as they act as class descriptors of successful partnerships. Finally, our data mining analyses revealed a difference in popular genre patterns in regional markets, in which local genres play a crucial role in musical success.

Overall, we contribute to the Hit Song Science field by advancing the knowledge on musical genres and collaborations as features in success-based models and making available a novel dataset with regional and temporal information. Such findings benefit not only scientists but also the music industry. In fact, in the past few years, several music actors have been using data insights to perform diagnostic analyses on the market and support business decisions. For instance, Instrumental is a data-driven British service that aims to use artificial intelligence and machine learning techniques to discover high potential talents and offer the most promising partnerships for independent artists.[5] From such a perspective, our work may contribute to music companies by enhancing such predictive and recommendation models with genre information. Therefore, both artists and record labels benefit from our findings, as they shed light on the science behind musical success and contribute to developing strategies to promote musical content across the world.

Acknowledgements. The first author thanks CNPq for his Masters Scholarship.

8. Research Outcomes

This Master thesis had its preliminary methodology and results published at the Workshop of Theses and Dissertations on Databases (WTDBD SBBD 2020) [Oliveira et al. 2020a] and was selected for publication at the Theses and Dissertations Contest of the Congress of the Brazilian Computer Society (CTD CSBC 2022) [Oliveira et al. 2022b]. Furthermore, this research contributed directly to other products, including:

- Study on collaboration profiles in success-based genre networks [Oliveira et al. 2020b] (*Best Paper Presentation Award at ISMIR 2020*);
- Application of the hot streak methodology in the US and Brazil [Barbosa et al. 2021, Seufitelli et al. 2022, Oliveira et al. 2023a] (*Best Paper Award at SBCM 2021*);
- MUHSIC and MUHSIC-BR, datasets with temporal information on musical success [Oliveira et al. 2021a, Oliveira et al. 2021b, Oliveira et al. 2022a];
- MGD, a dataset with genre success information [Oliveira et al. 2020c];
- Analyses on patterns in hit songs [Oliveira and Moro 2023a, Oliveira and Moro 2023b].

Byproducts. The knowledge acquired from this research has also contributed to:

- Teaching material that uses our datasets as basis, published as a book chapter and presented at CSBC 2021, SBBD 2021 and PPGC/UFF Winter School [Pimentel et al. 2021];
- An extensive survey on Hit Song Science [Seufitelli et al. 2023];
- Temporal comparative analysis on musical success [Silva et al. 2023];
- Opportunities to extend the concepts of the thesis in other MIR problems, such as hit song prediction [Silva et al. 2022a] (*Best Paper Runner-up at WebMedia 2022*) and mood analysis [Paula et al. 2022];
- Application of the thesis concepts in other contexts, including social coding [Moura et al. 2020, Oliveira et al. 2023b] and literature [Silva et al. 2021, Silva et al. 2022b] (*Best Paper Runner-up at BraSNAM 2021*);

References

Abel, F. et al. (2010). Analyzing the blogosphere for predicting the success of music and movie products. In *ASONAM*, pages 276–280, Odense, Denmark.

Askin, N. and Mauskapf, M. (2017). What makes popular culture popular? product features and optimal differentiation in music. *Amer. Sociolog. Rev.*, 82(5):910–944.

Barbosa, G. R. G., Melo, B. C., Oliveira, G. P., Silva, M. O., Seufitelli, D. B., and Moro, M. M. (2021). Hot Streaks in the Brazilian Music Market: A Comparison Between Physical and Digital Eras. In *SBCM*. SBC.

[5] About Instrumental: https://www.weareinstrumental.com/about

Bryan, N. J. and Wang, G. (2011). Musical influence network analysis and rank of sample-based music. In *ISMIR*, pages 329–334.

Cosimato, A. et al. (2019). The conundrum of success in music: Playing it or talking about it? *IEEE Access*, 7:123289–123298.

Dhanaraj, R. and Logan, B. (2005). Automatic prediction of hit songs. In *ISMIR*, pages 488–491, London, UK. ISMIR.

Garimella, K. and West, R. (2019). Hot streaks on social media. In *ICWSM*, pages 170–180. AAAI Press.

Janosov, M. et al. (2020). Success and luck in creative careers. *EPJ Data Sci.*, 9(1):9.

Keogh, E. J. and Pazzani, M. J. (2000). Scaling up dynamic time warping for data mining applications. In *KDD*, pages 285–289.

Klösgen, W. and Zytkow, J. M. (2002). *Handbook of data mining and knowledge discovery*. Oxford University Press, Inc.

Lundberg, S. M. and Lee, S. (2017). A unified approach to interpreting model predictions. In *NIPS*, pages 4765–4774.

Moura, A. F. C., Oliveira, G. P., Brandão, M. A., and Moro, M. M. (2020). Classification and persistence analysis of tie strength on github. In *Anais Est. WebMedia*, pages 41–44. SBC.

Oliveira, G. P., Barbosa, G. R. G., Melo, B. C., Botelho, J. E., Silva, M. O., Seufitelli, D. B., and Moro, M. M. (2022a). Musical Success in the United States and Brazil: Novel Datasets and Temporal Analyses. *JIDM*.

Oliveira, G. P., Barbosa, G. R. G., Melo, B. C., Silva, M. O., Seufitelli, D. B., Lacerda, A., and Moro, M. M. (2021a). MUHSIC: Music-oriented Hot Streak Information Collection. https://doi.org/10.5281/zenodo.4779003.

Oliveira, G. P., Barbosa, G. R. G., Melo, B. C., Silva, M. O., Seufitelli, D. B., and Moro, M. M. (2021b). MUHSIC: An open dataset with temporal musical success information. In *SBBD DSW*, pages 65–76. SBC.

Oliveira, G. P., Barbosa, G. R. G., Melo, B. C., Silva, M. O., Seufitelli, D. B., and Moro, M. M. (2023a). Hot streaks in the music industry: Identifying and characterizing above-average success periods in artists' careers. *Scientometrics*. [under review].

Oliveira, G. P., Lacerda, A., and Moro, M. M. (2020a). Musical genre analysis over dynamic success-based networks. In *SBBD WTDBD*. SBC.

Oliveira, G. P., Lacerda, A., and Moro, M. M. (2022b). Analyses of musical success based on time, genre and collaboration. In *Anais do XXXV CTD*. SBC.

Oliveira, G. P. and Moro, M. M. (2023a). Exceptional collaboration patterns in music genre networks. In *BraSNAM*. SBC.

Oliveira, G. P. and Moro, M. M. (2023b). Mining exceptional genre patterns on hit songs. In *KDMiLe*. SBC.

Oliveira, G. P., Moura, A. F. C., Batista, N. A., Brandao, M. A., Hora, A., and Moro, M. M. (2023b). How do developers collaborate? investigating github heterogeneous networks. *Software Quality Journal*, 31:211–241.

Oliveira, G. P., Silva, M. O., Seufitelli, D. B., Lacerda, A., and Moro, M. M. (2020b). Detecting collaboration profiles in success-based music genre networks. In *ISMIR*, pages 726–732.

Oliveira, G. P., Silva, M. O., Seufitelli, D. B., Lacerda, A., and Moro, M. M. (2020c). MGD: Music Genre Dataset. https://doi.org/10.5281/zenodo.4778563.

Paula, B. C. M., Oliveira, G. P., and Moro, M. M. (2022). Mood Analysis during the COVID-19 Pandemic in Brazil through Music. In *WebMedia CTIC*, pages 53–56, Porto Alegre, RS, Brasil. SBC.

Pimentel, J. F., Oliveira, G. P., Silva, M. O., Seufitelli, D. B., and Moro, M. M. (2021). Ciência de dados com reprodutibilidade usando jupyter. In *Jornada de Atualização em Informática 2021*, chapter 1, pages 13–62. SBC.

Ren, J. and Kauffman, R. J. (2017). Understanding music track popularity in a social network. In *Euro. Conf. Information Systems*, pages 374–388, Atlanta, USA. AIS.

Seufitelli, D. B., Oliveira, G. P., Silva, M. O., Barbosa, G. R. G., Melo, B. C., Botelho, J. E., Melo-Gomes, L., and Moro, M. M. (2022). From compact discs to streaming: A comparison of eras within the brazilian market. *Vortex Music Journal*, 10(1).

Seufitelli, D. B., Oliveira, G. P., Silva, M. O., Scofield, C., and Moro, M. M. (2023). Hit song science: A comprehensive survey and research directions. *JNMR*. [under review].

Shin, S. and Park, J. (2018). On-chart success dynamics of popular songs. *Advances in Complex Systems*, 21(3-4):1850008.

Silva, M. O. et al. (2019). Collaboration Profiles and Their Impact on Musical Success. In *Procs. of ACM/SIGAPP SAC*, pages 2070–2077, Limassol, Cyprus.

Silva, M. O. and Moro, M. M. (2019). Causality analysis between collaboration profiles and musical success. In *WebMedia*, pages 369–376. ACM.

Silva, M. O., Oliveira, G. P., Seufitelli, D. B., Lacerda, A., and Moro, M. M. (2022a). Collaboration as a driving factor for hit song classification. In *WebMedia*.

Silva, M. O., Oliveira, G. P., Seufitelli, D. B., and Moro, M. M. (2023). Temporal Success Analysis in Music Collaboration Networks: Brazilian versus Global Scenarios. *Vortex Music Journal*. [accepted for publication].

Silva, M. O., Scofield, C., de Melo-Gomes, L., Botelho, J. E., Oliveira, G. P., Seufitelli, D. B., and Moro, M. M. (2022b). Brazilian reading preferences in goodreads: Cross-state and cross-region analyses. *iSys*, 15(1).

Silva, M. O., Scofield, C., Oliveira, G. P., Seufitelli, D. B., and Moro, M. M. (2021). Exploring brazilian cultural identity through reading preferences. In *BraSNAM*, pages 115–126.

Sinatra, R. et al. (2016). Quantifying the evolution of individual scientific impact. *Science*, 354(6312).

Suh, M. M. et al. (2021). AI as social glue: Uncovering the roles of deep generative AI during social music composition. In *CHI*, pages 582:1–582:11. ACM.

Zaki, M. J. and Meira Jr., W. (2014). *Data Mining and Analysis: Fundamental Concepts and Algorithms*. Cambridge University Press.