Journal of Information and Data Management, 2025, 16:1, doi: 10.5753/jidm.2025.4676

Frequent Genre Mining on Hit and Viral Songs

Gabriel P. Oliveira [**Universidade Federal de Minas Gerais** | ***gabrielpoliveira@dcc.ufmg.br***]
Mirella M. Moro [**Universidade Federal de Minas Gerais** | ***mirella@dcc.ufmg.br***]

Department of Computer Science, Universidade Federal de Minas Gerais, Av. Antônio Carlos, 6627, Pampulha, Belo Horizonte, MG, 31270-010, Brazil.

Received: 6 July 2024 • **Published:** 18 March 2025

Abstract Music is a dynamic cultural industry that has produced large volumes of data since the beginning of streaming services. Understanding such data provides valuable insights into music consumption, and helps identifying emerging trends and fostering creativity within the music industry. Nowadays, combining different genres has become a common practice to promote new music and reach new audiences. Given the diversity of combinations between all genres, predictive and descriptive analyses are very challenging. This work aims to explore the relationship between genre combinations and music popularity by mining frequent patterns in hit and viral songs across global and regional markets. We extend previous work by incorporating viral songs into the analysis, thus strengthening the comparative analysis of musical popularity's interconnected facets. We use the Apriori algorithm to mine genre patterns and association rules that reveal how music genres combine with each other in each market. Our findings reveal significant differences in popular genres across regions and highlight the dynamic nature of genre-blending in modern music. In addition, we are able to use such patterns to identify and recommend promising genre combinations for such markets through the association rules.

Keywords: Musical Success, Viral Content, Musical Genres, Music Data Mining, Association Rules

1 Introduction

Music is not only one of the world's most important cultural industries but also one of the most dynamic. The volumes of music-related data available on the Web brings new challenges to the industry daily. Such extensive and complex data generated by music streaming platforms offer numerous research opportunities across various fields. Indeed, analyzing data on songs, their features, and the social interactions surrounding them can provide valuable insights into music consumption and listener behavior, as well as other industry-related aspects [Ghaffari *et al.*, 2024].

Understanding music-related data also helps identifying emerging trends and fostering creativity within the music industry. For example, as the music industry is complex and competitive, artists are encouraged to reinvent strategies to maintain their presence in the market and reach new audiences. A few years ago, music consumption was limited to radio and physical discs; now, the Internet and the popularization of streaming services enable fans all over the world to access an immense amount of songs from various artists. This new reality has facilitated the insertion of artists and regional genres in the global music scene, enhancing its diversity. Examples include the achievements of the Brazilian singer Anitta, whose song "Envolver" ranked #1 in the Global Spotify Top 200 Chart in 2022, and the establishment of K-pop as a popular genre worldwide.

In recent years, the use of social media and streaming platforms has also leveraged the viral phenomenon in the music domain. Besides being closely related, virality and success can be interpreted as two distinct faces of music popularity [Oliveira *et al.*, 2024]. In summary, **viral songs** relate to the fast dissemination of content on social platforms, whereas **music hits** (success) are more solid and associated with music consumption metrics (i.e., streams, radio airplay, and sales) [Guerini *et al.*, 2011; Seufitelli *et al.*, 2023b]. Nonetheless, viral trends frequently lead songs and artists to mainstream success. Examples include the song *Dreams* by Fleetwood Mac, a 1977 song that got back to the top of the charts in 2020 after a TikTok viral.[1]

When studying musical popularity, a very important issue is how to analyze the music market. Especially, analyzing the global performance is important, but not enough, as *regional markets* have their own features and behaviors regarding success and virality. Previous studies reveal that besides the globally established genres, regional genres are deeply consumed in their countries [Oliveira *et al.*, 2020]. In Brazil, the largest music market in Latin America, the five most listened songs in Spotify in 2022[2] include those from *sertanejo* and *funk carioca* – two of the most popular local genres.

Indeed, the genre perspective is essential when analyzing musical popularity. Within an ever evolving market, artists are no longer linked to one specific genre. Genre-blending has become common in music, and musicians often experiment with the mixture of genres in their songs. For example, Lil Nas X's "Old Town Road" (2019's #1 hit in the Billboard Year-End Hot 100) combines hip-hop and country, pop star Ariana Grande raps on her album *Positions*, and Bad Bunny navigates between reggaeton, EDM (electronic dance music), and trap in his songs.

[1]Billboard: `https://www.billboard.com/pro/fleetwood-mac-dreams-returns-hot-100/`

[2]CNN Brasil: `https://www.cnnbrasil.com.br/entretenimento/spotify-divulga-retrospectiva-de-2022-marilia-mendonca-e-a-mais-ouvida-no-brasil/`

Collaborating with artists from other genres is also a common approach to achieve greater popularity. For example, Lady Gaga's Grammy-nominated album *Chromatica*[3] has collaboration as one of its biggest strengths. The collaboration with Ariana Grande (*pop*), Elton John (*rock*), and Blackpink (*k-pop*) contributed to maintain her among the most prominent pop names, as well as introducing her to new audiences. More recently, the collaboration between Shakira and Bizarrap, entitled "Music Session Vol.53", took the record for the most-streamed Latin track in a single day in Spotify history on the day of its release in January 2023. It debuted in #1 in the Global chart and in several Spanish-speaking markets, including Argentina, Colombia, Chile, Mexico, and Spain. It also has made it to the top 10 in Luxembourg (#5), the United States (#6), and Switzerland (#7).

The great diversity of genres and the ever-growing tendency to blend them in songs highlight the dynamic and unpredictable nature of the music industry. Given the potential combinations between all such genres, predictive and descriptive analyses over the data available become challenging. As a practical application, record labels may uncover frequent genre combinations that achieve a higher level of success to plan future song releases. In such an interesting, challenging context, we aim to mine relevant patterns of musical genres in songs that have been successful (i.e., hit songs) and viral in both global and regional markets. In other words, we want to verify if there is a relationship between combining different musical genres and popularity. We do so by answering the following research questions (RQs):

RQ1. Compared to the global scenario, do regional markets present distinct patterns of frequent genre combinations in hit and viral songs?

RQ2. Is it possible to identify and recommend combinations of music genres that are promising and relevant to each market?

This work extends the paper presented at the 11th Symposium on Knowledge Discovery, Mining and Learning (KDMiLe 2023) [Oliveira and Moro, 2023]. As a novel contribution, we enhance our analysis by incorporating viral songs from both global and regional markets. Such an addition allows for identifying not only the emerging genre patterns in hit songs but also the most frequent combinations in viral songs. Therefore, the new results strengthen the comparative analysis between these two distinct yet interconnected facets of musical popularity by providing a better understanding of the dynamics of music trends worldwide.

Our findings reveal that there is indeed a difference in popular genres across regional markets, and we are able to use such patterns to identify and recommend promising genre combinations for such markets. The remainder of this paper is organized to answer all such questions. We first present relate work in Section 2 and introduce our methodology based on data mining techniques in Section 3. Then, Section 4 details the experimental evaluation and presents the results, and Section 5 discusses the results by comparing the findings from both hit and viral songs. Finally, we make our overall considerations in Section 6.

2 Related Work

Over the years, the number of artists and songs has increased considerably, as have studies on discovering the recipe for turning a song into a hit. Such studies define the area of *Hit Song Science* (HSS), which combines machine learning and data mining techniques with musicology and psychology concepts to verify whether popular songs share similar patterns. The concept of HSS was introduced in 2003 by Polyphonic HMI,[4] and since Dhanaraj and Logan [2005], different studies analyze the impact of acoustic and social characteristics on musical success [Calefato *et al*., 2018; Cosimato *et al*., 2019; Silva *et al*., 2022; Mayerl *et al*., 2023], some including genre information [Abel *et al*., 2010; Zangerle *et al*., 2019; Gienapp *et al*., 2021].

Over the years, several music-related datasets have also been proposed to allow research and analyses on HSS throughout distinct sets of features. However, most datasets consider data from a single music market (mostly the United States, i.e., the world's biggest music market), which may not represent the whole global scenario [Oliveira *et al*., 2020; Seufitelli *et al*., 2023a]. Therefore, recent datasets aim to cover broader perspectives. Focusing on the Brazilian Market, Bertoni and Lemos [2022] introduce four datasets with acoustic features from popular songs. In addition, MUHSIC-BR contains structured data on the temporal evolution of artists' careers [Oliveira *et al*., 2022]. Going further, MGD (and its extension, MGD+) uses data obtained from Spotify to provide, in addition to acoustic features and metadata, success collaboration networks on several music markets and also in the global scenario [Seufitelli *et al*., 2023a].

Although interconnected, success and virality represent two distinct facets of music popularity [Oliveira *et al*., 2024]. Existing research on music virality delves into viral marketing and listener behavior [Kahl and Albers, 2013; Fink *et al*., 2021], as well as how a song's presence on viral charts can predict its subsequent success [Araujo *et al*., 2019]. More recently, music has also been extensively shared in TikTok, which has emerged as a primary platform for content virality. Studies on TikTok explore personal motivations, behaviors [Compte and Klug, 2021], and content features that drive virality, focusing on content elements, the recommendation system, and the creator's profile [Ling *et al*., 2022].

Still, understanding musical aspects can be genre-dependent, which also reflects in musical success. For example, Ren and Kauffman [2017] aggregate genres in a musical construct vector (MCV) to summarize the acoustic content of a song in a predictive model. Then, Shin and Park [2018] consider genres to understand the life trajectory of songs in Gaon Chart, one of the main Korean

[3] *Chromatica* was nominated for Best Pop Vocal Album at the 63rd Grammys (2021).

[4] Polyphonic HMI, Hit Song Science: `https://polyphonichmi.blogspot.com/p/about-company.html`

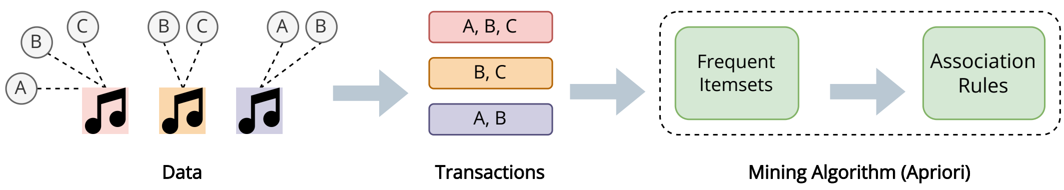

Figure 1. Methodology used to find frequent and exceptional patterns in hit and viral songs.

music rankings. Furthermore, Ordanini *et al.* [2018] use genre distance and other variables to show that artists who collaborate with artists from other genres are able to achieve a higher level of success in their songs. In addition, mining techniques have been used in the musical context on file indexing by style [Rompré *et al.*, 2017] and recognition of style hierarchies [Iloga *et al.*, 2018].

Studying musical success from a genre perspective may reveal important information on how artists from different communities try to achieve popularity. Our main contribution relies on advancing the understanding of the factors behind both musical success and virality. Specifically, we apply data mining techniques to uncover the most frequent and promising genre combinations in hit and viral songs. Considering regional markets makes this work more realistic, as each one behaves in its own distinct way. Thus, our results may enhance further analyses to guide future song releases.

3 Methodology

This section presents the methodology used to find both frequent and exceptional patterns in hit and viral songs (Figure 1). After defining virality and success (Section 3.1), we use a dataset with songs of global and regional markets (Section 3.2). Then, we model songs as transactions to find frequent itemsets and association rules for each market (Section 3.3). Finally, we use an itemset mining algorithm to find frequent genre combinations and three metrics to evaluate its association rules (Section 3.4).

3.1 Defining Virality and Success

Musical virality and success are not synonymous. Despite being closely connected, they are distinct facets of a song's popularity, a broader concept frequently associated with getting noticed by many people [Werber *et al.*, 2023]. We make such a separation following the music industry trend. For example, streaming services and specialized magazines started to differentiate viral from hit (i.e., successful) songs in their popularity charts. Therefore, distinguishing the two concepts is fundamental to understanding the dynamics of music consumption in the digital age.

The concept of **virality** is related to quickly spreading and disseminating content across various platforms and social networks [Guerini *et al.*, 2011, 2012]. For music, a viral song gains widespread attention when it is shared by thousands or even millions of users in a very short time span. Indeed, this concept is strictly related to social platforms, and it does not have a unique metric for it. For example, Guerini *et al.* [2011] define virality as the number of people who accessed a specific content in a given time interval, being associated with other measurable phenomena, including appreciation, buzz, and controversiality. In the music industry, specialized magazines and streaming services released virality charts that consider metrics such as the number of accesses and engagements (e.g., TikTok Billboard Top 50,[5] Spotify Viral 50,[6] YouTube Trending Videos[7]).

In contrast, musical **success** is associated with the music consumption itself, but it also does not have a unique metric or definition. Still, according to Seufitelli *et al.* [2023b], the most used success definitions in the field of Hit Song Science (HSS) can be divided into three main perspectives: (*i*) top-chart, which are rankings of songs based mainly on sales, media airplay, and/or online streaming (e.g., Billboard Hot 100,[8] Official UK Singles Chart,[9] Crowley Top 100 Brasil[10]); (*ii*) economy indicators, such as the number of single/album sales; and (*iii*) engagement, which considers the social dimension of success, including the number of views or likes on social platforms.

To analyze the dynamics of music popularity in current times, we consider the top-chart perspective for measuring both virality and success. Therefore, we consider as viral all songs that have entered a viral chart, whereas hits are those songs that have made it into a distinct success chart (e.g., the most streamed songs). We do so regardless of their position in the charts, i.e., songs ranked first and last are equally considered viral/hit.

3.2 Data

In this work, we use the Music Genre Dataset (MGD) [Oliveira *et al.*, 2020; Seufitelli *et al.*, 2023a] as source of Spotify data from 2017 to 2019.[11] Spotify provides charts of hit and viral songs for each country and territory, as well as an aggregated global chart. We consider as **hit songs** the ones that are present in the Top 200 Charts, which contain the most streamed songs in a specific market. Furthermore, **viral songs** are those that have reached the Viral 50 Charts, which feature the songs gaining the most attention on the

[5] `https://www.billboard.com/charts/tiktok-billboard-top-50/`
[6] `https://charts.spotify.com/`
[7] `https://charts.youtube.com/`
[8] `https://www.billboard.com/charts/hot-100/`
[9] `https://www.officialcharts.com/charts/singles-chart/`
[10] `https://charts.crowley.com.br/index.html`
[11] Because of significant changes in the Spotify Charts platform, freely downloading the charts is no longer possible. Therefore, we were not able to perform our analysis in more recent data.

Table 1. Dataset main numbers.

	Hit	Viral
Songs	13,880	29,546
Artists	3,612	20,341
Genres	896	1,710
Super-genres	162	218

platform based on increases in plays, shares, and the number of people who have recently discovered them.[12]

The dataset considers the global charts and eight of the top 10 music markets according to IFPI in 2019:[13] United States (#1), Japan (#2), United Kingdom (#3), Germany (#4), France (#5), Canada (#8), Australia (#9), and Brazil (#10).[14] MGD provides the songs that entered in the charts for each market and year, as well as acoustic features that describe such songs. In addition, it provides relevant information on the artists who interpret the songs, including their genre list.

Regarding artists' genres, the MGD buiding process contains a preprocessing phase to generate the final genre list. This is because Spotify allows inserting very specific genres (or subgenres) for each artist. For example, Shakira is assigned to both *colombian pop* and *pop*, which may be described only by *pop*. Then, the dataset maps all specific genres to more embracing *super-genres*. Table 1 presents the main numbers of the final dataset.

3.3 Modeling Songs as Transactions

Discovering frequent itemsets and association rules requires a transactional dataset. In such modeling, the dataset instances (i.e., transactions) are composed of a list of items. Here, we model an individual transactional dataset for each market and year to find the most frequent genre combinations and association rules. We define each hit and viral song as a single transaction in which the items are the musical genres of the artists who sing it. If a song has more than one artist and they all share a common genre, this genre appears only once in the transaction. Therefore, we are not interested in the number of repeated genres in a song, but in the diversity of different genres that compose it.

For example, consider the remix version of *Despacito* by Luis Fonsi, Daddy Yankee, and Justin Bieber, which spent 16 consecutive weeks in the #1 position on Billboard Hot 100 in 2017. The transaction correspondent to this song comprises the genres of all three artists: *latin*, *tropical*, *pop*, *reggaeton*, and *hip hop*. Besides the fact that *latin* and *tropical* are shared by Fonsi and Yankee and *pop* is shared by Fonsi and Bieber, each genre is counted once. Thus, the final transaction for *Despacito* is represented by the tuple `(hip hop, latin, tropical, pop, reggaeton)`.[15]

3.4 Mining and Metrics

We focus on finding the most frequent genre combinations by applying a Frequent Itemset Mining (FIM) method from the set of hit songs in each musical market from MGD. We do so by running the Python implementation of Apriori,[16] one of the most popular FIM algorithms [Agrawal *et al.*, 1994]. We perform both temporal and regional analyses, running the algorithm separately for each market and year (2017 to 2019). Following the methodology of Section 3.3, we define the transactions of the FIM task as songs (hit and viral), whose items are the musical genres of each artist who sings them. Each song appears once in the dataset, as our goal is not to evaluate their time within the charts.

Once mined, frequent itemsets can be used to generate Association Rules (AR). An AR is represented by the expression $X \rightarrow Y$ and composed of an antecedent X and a consequent Y, two disjoint itemsets. An AR should not be interpreted as a sign of causality but of co-occurrence between items. Indeed, association rules allow to discover how itemsets are related. To evaluate our frequent itemsets and association rules, we use three classic data mining metrics: relative support, confidence, and lift (definitions are further detailed in [Zaki and Meira Jr., 2014]).

Relative Support. Denoted by $rsup(X)$, it informs the frequency (or the empirical probability) in which an itemset X appears on the transactions on a scale from 0 to 1. Then, a relative support of 1 means that an itemset occurs in all transactions.

Confidence. The rule confidence informs the probability of a consequent Y occurring in a transaction given the occurrence of an antecedent X. In other words, it is the frequency in which Y occurs in transactions containing X.

$$conf(X \rightarrow Y) = P(Y|X) = \frac{P(XY)}{P(X)} = \frac{rsup(XY)}{rsup(X)}$$

Lift. It is the ratio between the joint probability of X and Y co-occurring and the probability of these sets being independent. It may be used as a measure of surprise within a rule. Therefore, lift shows how much more frequently the consequent Y becomes after the occurrence of the antecedent X. Such a metric is symmetric, and values below 1 mean that the rule occurs less than expected, whereas values above 1 indicate the opposite.

$$lift(X \rightarrow Y) = \frac{P(XY)}{P(X) \cdot P(Y)} = \frac{conf(X \rightarrow Y)}{rsup(Y)}$$

4 Results

This section presents the results and discussions for each analysis carried out to answer our research questions: frequent pattern mining (*RQ1*, Section 4.1) and association rules (*RQ2*, Section 4.2).

4.1 Genre Frequent Patterns

Discovering changes in genre preferences shows the dynamic nature of the music market. As an important

[12] Spotify: `https://support.spotify.com/us/artists/article/charts/`

[13] IFPI Global Music Report 2019: `https://gmr.ifpi.org/`

[14] Data from South Korea (#6) and China (#7) was not available for such a period.

[15] Sorted by alphabetical order.

[16] PyFIM: `https://borgelt.net/pyfim.html`

Table 2. Top 5 most frequent patterns in global and English-speaking markets (2019).

	Hit		Viral	
Market	**Pattern**	**Support**	**Pattern**	**Support**
Global	(dance pop, pop)	0.271	(hip hop, trap)	0.145
	(latin, reggaeton)	0.173	(reggaeton, latin)	0.127
	(hip hop, trap)	0.172	(latin, trap)	0.125
	(rap, hip hop)	0.168	(reggaeton, trap)	0.119
	(rap, trap)	0.151	(rap, hip hop)	0.117
Australia	(dance pop, pop)	0.294	(trap, hip hop)	0.130
	(rap, hip hop)	0.162	(rap, hip hop)	0.122
	(electropop, pop)	0.145	(trap, rap)	0.111
	(rap, pop rap)	0.145	(trap, rap, hip hop)	0.084
	(pop rap, hip hop)	0.131	(pop rap, hip hop)	0.062
Canada	(hip hop, rap)	0.273	(rap, hip hop)	0.199
	(trap, rap)	0.255	(trap, hip hop)	0.184
	(dance pop, pop)	0.253	(trap, rap)	0.180
	(pop rap, rap)	0.252	(trap, rap, hip hop)	0.142
	(hip hop, pop rap)	0.225	(hip hop, pop)	0.089
UK	(dance pop, pop)	0.285	(rap, hip hop)	0.129
	(rap, hip hop)	0.159	(trap, hip hop)	0.094
	(tropical house, pop)	0.133	(trap, rap)	0.094
	(tropical house, dance pop)	0.127	(trap, rap, hip hop)	0.069
	(tropical house, dance pop, pop)	0.125	(grime, hip hop)	0.067
USA	(hip hop, rap)	0.305	(hip hop, rap)	0.193
	(trap, rap)	0.289	(hip hop, trap)	0.192
	(pop rap, rap)	0.261	(rap, trap)	0.185
	(trap, hip hop)	0.246	(hip hop, rap, trap)	0.138
	(pop rap, hip hop)	0.230	(trap, pop)	0.101

cultural artifact, the way music is consumed in different countries may be influenced by language, demographics, and other social aspects. In addition, musicians are naturally venturing into new domains and working outside of the style they had initially emerged from, resulting in a massive variety of new songs and musical tastes. Therefore, in this section, we answer *RQ1* by investigating whether genre combination varies at a country level. That is, we analyze if the association of distinct musical genres in hit songs has specific patterns in global and regional markets.

As language is crucial for listening to music, we divide our eight regional markets into two distinct groups: English and non-English speaking countries. The former includes Australia, Canada, the United Kingdom, and the United States, while the latter comprises Brazil, France, Germany, and Japan. 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. Here, we present the results for selected markets and years for readability purposes.

English-speaking markets. Table 2 presents the five most frequent genre patterns in hit and viral songs for global and English-speaking markets[17] in 2019. Itemsets are sorted by their relative support value, i.e., their frequency. Regarding the global scenario, both hit and viral songs have a strong presence of mainstream genres such as *pop*, *hip hop*, and *rap*. These genres include regional versions of themselves (e.g., *Chicago rap* is within *rap*), which may contribute for their high support values. Besides, the combination of the regional genres *latin* and *reggaeton* appear in 17.3% of all global hit songs and in 12.7% of virals showing their popularization across the world. The Latin music expansion began in the early 2000s with artists like Shakira and Ricky Martin, reaching greater popularity in the late 2010s, led by artists such as Bad Bunny, J Balvin, and Karol G.

By analyzing the English-speaking countries individually, we note a high similarity in the popular genre combinations. For instance, the union of *hip hop* and *rap*, which is present in 30.5% of hit songs in the United States, is also relevant in Australia (16.2%), Canada (27.2%) and the United Kingdom (15.9%). All such countries present a strong similarity to the global market, as they share several cultural aspects, including language (English is the most widely spoken language worldwide in terms of number of countries where it is official – 59 countries in all continents).[18]

In all such markets, viral songs present a similar behavior. Specifically, the combinations of *rap*, *hip hop*, *trap*, and *pop* consistently appear among the most frequent patterns. The novelty here lies in the UK, which presents the pattern `(grime,`[19] `hip hop)` among the most frequent. However, the support values of all such itemsets are significantly lower than those observed in hit songs. This indicates that such genre combinations are not as prevalent in viral songs as in successful ones, suggesting a greater variability in the musical genre combinations present in viral songs.

Non-English-speaking markets. On the other hand, the analysis of frequent genre patterns for non-English

[17] We consider Canada as an English-speaking country due to the majority of its population being English speakers, despite its bilingual status.

[18] World Atlas (Access in July, 2024): `https://www.worldatlas.com/articles/the-most-popular-official-languages-of-the-world.html`

[19] Grime is a genre of electronic dance music that emerged in the early 2000s in London, characterized by its aggressive beats, rapid tempos, and gritty lyrical content reflecting urban life.

Table 3. Top 5 frequent patterns in global and non-English speaking markets (2019).

	Hit		Viral	
Market	**Pattern**	**Support**	**Pattern**	**Support**
Global	(dance pop, pop)	0.271	(hip hop, trap)	0.145
	(latin, reggaeton)	0.173	(reggaeton, latin)	0.127
	(hip hop, trap)	0.172	(latin, trap)	0.125
	(rap, hip hop)	0.168	(reggaeton, trap)	0.119
	(rap, trap)	0.151	(rap, hip hop)	0.117
Brazil	(brazilian funk, pop)	0.177	(arrocha, sertanejo)	0.116
	(electro, brazilian funk)	0.102	(hip hop, trap)	0.073
	(sertanejo, brazilian funk)	0.097	(brazilian funk, pop)	0.063
	(electro, pop)	0.080	(funk pop, brazilian funk)	0.056
	(trap, hip hop)	0.064	(brega funk, brazilian funk)	0.056
France	(hip hop, pop)	0.584	(hip hop, pop)	0.260
	(rap, hip hop)	0.449	(rap, hip hop)	0.228
	(rap, pop)	0.423	(rap, pop)	0.201
	(rap, hip hop, pop)	0.393	(rap, hip hop, pop)	0.185
	(francoton, pop)	0.174	(r&b, pop)	0.124
Germany	(dance pop, pop)	0.162	(rap, hip hop)	0.120
	(rap, hip hop)	0.158	(hip hop, pop)	0.070
	(hip hop, pop)	0.130	(trap, hip hop)	0.063
	(tropical house, pop)	0.105	(trap, rap)	0.056
	(tropical house, dance pop)	0.087	(rap, pop)	0.052
Japan	(j-rock, j-pop)	0.283	(j-rock, j-pop)	0.143
	(other, j-pop)	0.140	(anime, j-pop)	0.108
	(anime, j-pop)	0.138	(punk rock, j-pop)	0.049
	(dance pop, pop)	0.133	(j-punk, pop punk, punk rock, j-rock, j-pop)	0.041
	(r&b, j-pop)	0.108	(other, j-pop)	0.033

speaking countries reveals a strong regional component in most countries. Table 3 presents the five most frequent genre associations in 2019 for four countries: Brazil, France, Germany, and Japan. All such countries have patterns with regional rhythms, such as *francoton* in France, and *brazilian funk* and *sertanejo* in Brazil. However, Japan stands out in this regard, as all five patterns have regional styles. The main genres in such a market include *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 four 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.

Regarding viral songs for such markets, Brazil and Japan have distinct regional genres among the most frequent patterns. For example, the combination of *arrocha* and *sertanejo* is the most frequent in viral songs in Brazil in 2019, with such genres co-occurring in 11.6% of viral songs. Whereas *sertanejo* is a genre that is already well established and popular in the country, *arrocha*[20] is a more recent genre that became popular in the 2000s and is achieving national popularity through its partnerships and influences.[21]

Furthermore, viral song patterns in European markets (France and Germany) are similar to those in English-speaking markets, with strong presence of *rap*, *hip hop*, and *pop* in the most frequent patterns. This result can be explained by the cultural proximity between such countries, which was boosted by globalization and the emergence of streaming platforms as the main format for music consumption. In fact, globalization has facilitated cultural exchange, allowing musical trends to spread quickly from one country to another, resulting in a convergence of musical tastes.

We are now able to answer RQ1 – *Compared to the global scenario, do regional markets present distinct patterns of frequent genre combinations in hit and viral songs*? In short, yes. Whereas most markets have similarities with the global scenario, especially regarding genres such as *pop*, *rap*, and *hip hop*, regional markets still have their particularities. This is more notable in the non-English-speaking markets such as Brazil and Japan, which have a strong presence of regional genres (e.g., *sertanejo*, *brazilian funk*, *j-pop*, *anime*). Indeed, this blend of global and local influences results in a rich diversity of musical expressions that distinguishes regional markets from the global scenario.

4.2 Promising Genre Associations

Using the data mining framework offers a wide range of possibilities to perform descriptive analyses in datasets [Fontes *et al.*, 2019; Amaral and de Sousa, 2020; Melo *et al.*, 2021]. For instance, the frequent genre patterns mined in Section 4.1 can be used to uncover association rules, which inform how items (i.e., music genres) are associated with each other. In this section, we answer *RQ2* by using such rules to detect and recommend outstanding genre associations. We define promising rules according to their lift value, which measures their level of surprise. We

[20] *Arrocha* is a Brazilian musical genre originated in the state of Bahia, with influences of *bolero* and *brega* music. It is characterized by songs performed mainly with keyboards, electronic drums, and saxophone.

[21] Estadão: `https://expresso.estadao.com.br/naperifa/genero-musical-arrocha-revela-talentos-do-interior-da-bahia-para-o-brasil/`

Table 4. Association rules in global and regional markets sorted by lift value (2019).

	Hit			Viral		
Market	**Rule**	**Lift**	**Confidence**	**Rule**	**Lift**	**Confidence**
Global	(latin, reggaeton) → tropical	7.922	0.468	(reggaeton, trap) → latin	6.951	0.958
	(latin) → tropical	7.821	0.462	(latin) → reggaeton	6.671	0.920
	(reggaeton) → tropical	7.722	0.456	(reggaeton) → latin	6.671	0.920
Australia	(tropical house) → house	7.655	0.342	(trap, rap) → pop rap	6.152	0.428
	(tropical house, pop) → house	7.173	0.321	(rap, hip hop) → pop rap	6.128	0.426
	(tropical house, pop) → electro	7.111	0.670	(trap, hip hop) → pop rap	5.409	0.376
Brazil	(hip hop) → trap	6.187	0.434	(hip hop) → pop rap	5.864	0.388
	(brazilian funk, pop) → pagode baiano	5.473	0.425	(trap) → pop rap	5.083	0.337
	(hip hop) → pop rap	5.235	0.303	(hip hop) → trap	4.383	0.591
Canada	(r&b) → soul	7.485	0.226	(trap, rap, hip hop) → pop rap	4.927	0.467
	(dance pop) → tropical house	3.214	0.243	(trap, rap) → pop rap	3.961	0.375
	(dance pop, pop) → tropical house	3.160	0.239	(trap, hip hop) → pop rap	3.864	0.366
France	(rap, pop) → hip hop	1.301	0.900	(trap) → latin	8.254	0.628
	(rap, pop) → francoton	1.263	0.325	(trap) → reggaeton	8.011	0.611
	(hip hop, pop) → rap	1.259	0.796	(trap) → pop rap	5.989	0.249
Germany	(dance pop) → tropical house	5.909	0.400	(rap, hip hop) → pop rap	5.659	0.226
	(dance pop) → electro	5.908	0.338	(rap, hip hop) → trap	3.585	0.353
	(dance pop, pop) → tropical house	5.824	0.394	(rap) → trap	2.901	0.286
Japan	(r&b) → j-rap	8.067	0.228	(anime, j-pop) → visual kei	9.191	0.272
	(dance pop) → electro	4.348	0.283	(anime) → visual kei	8.020	0.237
	(dance pop, pop) → electro	4.284	0.279	(j-rock, j-pop) → j-punk	7.015	0.288
UK	(rock) → indie rock	8.370	0.364	(trap) → pop rap	5.247	0.291
	(rock) → indie	6.216	0.231	(rap, hip hop) → pop rap	5.155	0.286
	(pop rap, hip hop) → trap	5.682	0.660	(rap, hip hop) → trap	3.979	0.537
USA	(pop rap, pop, rap) → r&b	2.990	0.291	(trap, pop) → escape room	7.664	0.252
	(pop, rap) → r&b	2.888	0.281	(trap, pop) → reggaeton	5.559	0.399
	(hip hop, pop) → r&b	2.878	0.280	(trap, pop) → latin	5.462	0.387

then look for rules with high lift values to find the most promising genre associations. Here, we still perform an individual analysis for each market and year, using the Apriori algorithm to find the relevant rules.

Table 4 presents the three most promising rules for each market in 2019. We sort the rules by their lift values, but we also present the confidence value to enrich our analyses. The results for the global market reveal the strong association of regional genres such as *latin*, *reggaeton*, and *tropical*. With the lift value of the first rule, we can affirm the occurrence of *tropical* when *latin* and *reggaeton* co-occur in a hit song is almost eight times the expected. Such a result indicates that adding *tropical* in songs containing *latin* and *reggaeton* may increase in up to 7.922 times the chances of reaching the Top 200 chart. Besides, the rule confidence informs that *tropical* is present in 46.8% of the transactions (i.e., hit songs) that contain *latin* and *reggaeton*. Indeed, Latin musical genres had a boom in 2019, following the continuous growing trend observed in the late 2010s. Despite not achieving the #1 position as Luis Fonsi and Daddy Yankee did in 2017 with *Despacito*, artists such as Karol G, Bad Bunny, Ozuna, and J Balvin expressed the power of such genres, as they managed to put two or more songs in the charts.

We also note the presence of local genres in outstanding association rules, mainly in non-English speaking countries. For instance, *francoton* is associated with *rap* and *pop* in France, whereas the probability of *j-rap* occurring is eight times higher given the presence of *r&b* in Japanese hit songs. In Brazil, the genre *pagode baiano* (consequent) appears on 42.5% of the songs containing *brazilian funk* and *pop* (antecedent). In addition, the occurrence of the antecedent of such a rule increases more than five times the chances of the consequent in Brazilian hit songs. Thus, combining such genres considerably increases the chances of a song reaching the Brazilian charts. The singer-songwriter Anitta is an example of such an effect, as her music style list includes all three genres aforementioned. She is one of the most popular artists in the country, and her singles *Onda Diferente* (with Ludmilla) and *Combatchy* (with Lexa, Luísa Sonza, and MC Rebecca) contributed to the high relevance of associating such genres.

Regarding viral songs, the results are similar for the global market in the most promising genre associations. That is, there is a strong association between regional genres that achieved global popularity (i.e., *latin* and *reggaeton*). In addition, the rule with the highest lift value informs that combining *latin* (consequent) with both *reggaeton* and *trap* (antecedent) increases almost seven times the chances of a song being present in Spotify Viral 50 Charts. The confidence value means of such a rule that such an association happens in 95.8% of the songs containing the antecedent genres.

Moreover, *reggaeton* and *latin* are also present in outstanding association rules for viral songs in France and the United States. In both countries, such genres are associated with *trap* and *pop*/*pop rap*. The most promising genre associations for all other countries are mostly composed of *trap*, *rap*, and *hip hop*. This finding highlights

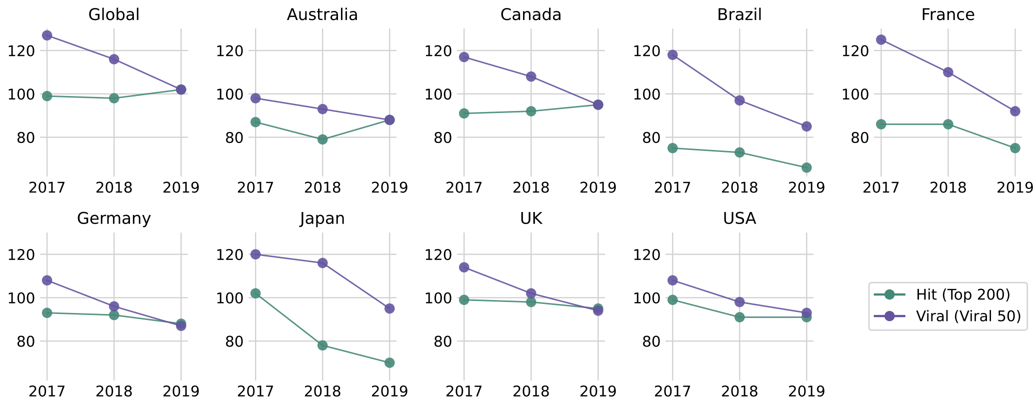

Figure 2. Number of distinct genres in hit and viral charts per year.

the strength of these genres in such countries, showing how they propel users to share songs and aid to their viral spread.

The exception is Japan, which only presents regional genres in their most promising association rules. Taking the first rule as an example, the occurrence of both *anime* and *j-pop* (antecedent) in the same song increases more than five times the chances of also having *visual kei* (consequent). In other words, the probability of having a song that combines the antecedent with the consequent is over five times higher than their individual probabilities. An example of such a combination is the song *Red Swan* by Yoshiki (*j-pop*) featuring HYDE (*anime*, *visual kei*), which was the opening theme song of the anime series *Attack on Titan*. Besides being on Spotify Viral 50, the song also reached Oricon and Billboard Japan Hot 100 Charts, confirming its popularity.

Overall, association rules are a powerful tool to understand musical success, as they reveal the level of combination between musical genres in global and regional markets. Similar to the previous sections, local genres play a fundamental role in regional markets, reinforcing their distinct cultural identities. Regarding RQ2 (*Is it possible to identify and recommend combinations of music genres that are promising and relevant to each market?*), we can affirm that using lift values to evaluate rules allows recommending promising genre combinations based on their high association level. Such an approach provides considerable benefits to artists, as they can plan their subsequent releases by choosing artists from genres with a high level of association with their own to collaborate. In addition, record labels may use our findings to diversify their set of artists and promote collaborations with high potential of success between them. Indeed, music is a dynamic and unpredictable industry, but this strategy may help guide artists and record labels to develop approaches to achieve success and increase their numbers.

5 Discussion

Despite sharing similarities, the frequent patterns and association rules found are not the same among hits and virals, even when considering the same market and year. Furthermore, the genres present in each set also differ, suggesting that users' music consumption and sharing behavior (i.e., the processes that lead to success and virality) follow different patterns. Whereas hits tend to follow more established and predictable genre combinations (e.g., *brazilian funk* and *pop* in Brazil, *rap* and *hip hop* in English-speaking markets), viral songs often explore more innovative and varied combinations. Examples include combining *arrocha* and *sertanejo* in Brazil, and *reggaeton* and *trap* in the Global scenario.

For association rules, the scenario is distinct, as hits present more diverse rules. They often combine regional with more established and globally popular genres, such as associating *francoton* with *hip hop* and *pop* in France. In contrast, most of the outstanding association rules for viral songs contain variations of the combination between *rap*, *hip hop*, and *trap*. While these genres are present in hit songs, they do not dominate as strongly as they do in viral tracks. An exception to this trend is Japan, where all the association rules prominently feature regional genres.

Furthermore, the main difference between hits and virals is the lower support values of the frequent patterns found in viral songs (see Tables 2 and 3). Such a phenomenon may indicate a greater diversity of genres within the analyzed set of songs. In other words, if the most frequent itemsets have low support, there may be several other music genres in the other songs. Figure 2 illustrates such a hypothesis, showing that viral charts mostly present a wider range of distinct musical genres in all markets and years.

6 Conclusion

In this work, we investigate the relationship between the combination of different musical genres and popularity from a data mining perspective. Using data from the Music Genre Dataset, which contains Spotify chart information from several markets, we perform descriptive analyses to identify frequent genre combinations and association rules in hit and viral songs. We conduct temporal analyses for both global and regional markets, i.e., we run the data mining algorithms individually for each market and year (2017 to 2019). Such an approach is helpful in revealing the evolution of musical tastes over time and showing how cultural aspects shape local music markets.

After modeling hit songs as transactions in which the items are their musical genres, our results reveal there is indeed a difference in popular genre patterns in regional markets, mainly in non-English speaking countries. In addition, we mined association rules to recommend promising genre combinations based on their level of surprise. Again, we found that local genres play a fundamental role in regional markets as they are included in most of the relevant associations.

Also, analyzing the viral songs and comparing them to hits allow a better understanding of the distinction between the processes behind the two categories. Indeed, both hits and virals present distinct patterns of genre combinations and association rules, which is in line with previous work that considers virality and success as two interconnected sides of musical popularity. However, more in-depth analyses are required to understand the factors behind such processes and how they are related to music consumption.

Finally, 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.

Limitations and Future Work. One limitation of this work is not considering how many times a hit/viral song appears in the charts when building the transactional dataset. Such a strategy would emphasize the genre frequency but, on the other hand, the distinctness of hit songs would be lost. In addition, we do not use the position in the charts in our analyses. Thus, songs that reached the Top 10 are treated equally to songs from the bottom of the charts. Further experiments must evaluate the impact of such information on genre popularity compared to the analyses already performed here. Moreover, we plan to perform our analyses on more recent data from other data sources to continue evaluating how music has evolved regarding genre association. Finally, other metrics can be used to enhance the frequent pattern and rule assessment as well. For example, conviction is a metric for the expected error in a rule, measuring its strength regarding the complement of the consequent.

Acknowledgements

The authors would like to thank Iago A. D. Vaz for his contribution in the early stages of this work.

Funding

This research was funded by CAPES, CNPq, and FAPEMIG, Brazil.

Competing interests

The authors declare that they have no competing interests.

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. DOI: 10.1109/ASONAM.2010.50.

Agrawal, R. *et al*. (1994). Fast algorithms for mining association rules. In *VLDB*, volume 1215, pages 487–499.

Amaral, T. and de Sousa, E. P. M. (2020). Mining temporal exception rules from multivariate time series using a new support measure. *J. Inf. Data Manag.*, 11(3):228–246. DOI: 10.5753/jidm.2020.2020.

Araujo, C. V. S. *et al*. (2019). Predicting music popularity on streaming platforms. In *SBCM*, pages 141–148, São João del-Rei, Brazil. SBC. DOI: 10.5753/sbcm.2019.10436.

Bertoni, A. A. and Lemos, R. P. (2022). Using Musical and Statistical Analysis of the Predominant Melody of the Voice to Create datasets from a Database of Popular Brazilian Hit Songs. *J. Inf. Data Manag.*, 13(1). DOI: 10.5753/jidm.2022.2336.

Calefato, F. *et al*. (2018). Collaboration success factors in an online music community. In *ACM GROUP*, Sanibel Island, USA. DOI: 10.1145/3148330.3148346.

Compte, D. L. and Klug, D. (2021). "it's viral!" – A study of the behaviors, practices, and motivations of tiktok users and social activism. In *CSCW Companion*, pages 108–111, Online. ACM. DOI: 10.1145/3462204.3481741.

Cosimato, A. *et al*. (2019). The conundrum of success in music: Playing it or talking about it? *IEEE Access*, 7:123289–123298. DOI: 10.1109/ACCESS.2019.2937743.

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

Fink, L. K. *et al*. (2021). Viral tunes: changes in musical behaviours and interest in coronamusic predict socio-emotional coping during COVID-19 lockdown. *Humanities and Social Sciences Communications*, 8(1). DOI: 10.1057/s41599-021-00858-y.

Fontes, S. G. *et al*. (2019). Association rules mining applied in the animal movement exploratory analysis. In *KDMiLe*, pages 1–8. SBC. DOI: 10.5753/kdmile.2019.8782.

Ghaffari, M., Khan, G. F., Singh, S. P., and Ferwerda, B. (2024). The impact of COVID-19 on online music listening behaviors in light of listeners' social interactions. *Multim. Tools Appl.*, 83(5):13197–13239. DOI: 10.1007/s11042-023-16079-1.

Gienapp, L. *et al*. (2021). Topological properties of music collaboration networks: The case of jazz and hip hop. *Digit. Humanit. Q.*, 15(1). DOI: 10.5753/kdmile.2020.11971.

Guerini, M., Pepe, A., and Lepri, B. (2012). Do linguistic style and readability of scientific abstracts affect their virality? In *ICWSM*. The AAAI Press.

Guerini, M. *et al*. (2011). Exploring text virality in social networks. In *ICWSM*, pages 506–509, Barcelona, Spain. The AAAI Press. DOI: 10.1609/icwsm.v5i1.14169.

Iloga, S. *et al*. (2018). A sequential pattern mining approach to design taxonomies for hierarchical music genre recognition. *Pattern Anal. Appl.*, 21(2):363–380. DOI: 10.1007/s10044-016-0582-7.

Kahl, C. and Albers, A. (2013). How to unleash the virus - social networks as a host for viral music marketing. In *CBI*, pages 47–54, Vienna, Austria. IEEE Computer Society. DOI: 10.1109/CBI.2013.16.

Ling, C. *et al*. (2022). Slapping cats, bopping heads, and oreo shakes: Understanding indicators of virality in tiktok short videos. In *WebSci*, pages 164–173, Barcelona, Spain. ACM. DOI: 10.1145/3501247.3531551.

Mayerl, M. *et al*. (2023). Pairwise learning to rank for hit song prediction. In *BTW*, volume P-331 of *LNI*, pages 555–565. Gesellschaft für Informatik e.V.. DOI: 10.18420/BTW2023-26.

Melo, E. *et al*. (2021). Combining Data Mining Techniques to Analyse Factors Associated with Allocation of Socioeconomic Resources at IFMG. In *KDMiLe*, pages 89–96. SBC. DOI: 10.5753/kdmile.2021.17465.

Oliveira, G. P., Barbosa, G. R. G., Melo, B. C., Botelho, J. E., Silva, M. O., Seufitelli, D. B., and Moro, M. M. (2022). Musical Success in the United States and Brazil: Novel Datasets and Temporal Analyses. *J. Inf. Data Manag.*, 13(1). DOI: 10.5753/jidm.2022.2350.

Oliveira, G. P., da Silva, A. P. C., and Moro, M. M. (2024). What makes a viral song? Unraveling music virality factors. In *WebSci*, pages 181–190, New York, NY, USA. ACM. DOI: 10.1145/3614419.3644011.

Oliveira, G. P. and Moro, M. M. (2023). Mining exceptional genre patterns on hit songs. In *KDMiLe*, pages 97–104. SBC. DOI: 10.5753/kdmile.2023.232412.

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

Ordanini, A. *et al*. (2018). The featuring phenomenon in music: how combining artists of different genres increases a song's popularity. *Market. Letters*, 29:485–499. DOI: 10.1007/s11002-018-9476-3.

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

Rompré, L. *et al*. (2017). Using association rules mining for retrieving genre-specific music files. In *FLAIRS Conference*, pages 706–711. AAAI Press.

Seufitelli, D. B., Oliveira, G. P., Silva, M. O., and Moro, M. M. (2023a). MGD+: An Enhanced Music Genre Dataset with Success-based Networks. In *Proceedings of the 5th Dataset Showcase Workshop*, pages 36–47, Porto Alegre, RS, Brasil. SBC. DOI: 10.5753/dsw.2023.233826.

Seufitelli, D. B., Oliveira, G. P., Silva, M. O., Scofield, C., and Moro, M. M. (2023b). Hit song science: a comprehensive survey and research directions. *J. New Music Res.*, 52(1):41–72. DOI: 10.1080/09298215.2023.2282999.

Shin, S. and Park, J. (2018). On-chart success dynamics of popular songs. *Adv. in Comp. Systems*, 21(3-4):1850008. DOI: 10.1142/S021952591850008X.

Silva, M. O., Oliveira, G. P., Seufitelli, D. B., Lacerda, A., and Moro, M. M. (2022). Collaboration as a driving factor for hit song classification. In *WebMedia*, pages 66–74. ACM. DOI: 10.1145/3539637.3556993.

Werber, N., Stein, D., Döring, J., Albrecht-Birkner, V., Gerlitz, C., Hecken, T., Paßmann, J., Schäfer, J., Schubert, C., and Venus, J. (2023). Getting Noticed by Many: On the Transformations of the Popular. *Arts*, 12(1). DOI: 10.3390/arts12010039.

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

Zangerle, E. *et al*. (2019). Hit song prediction: Leveraging low- and high-level audio features. In *ISMIR*, pages 319–326.