Exceptional Collaboration Patterns in Music Genre Networks

Gabriel P. Oliveira, Mirella M. Moro

[1] Universidade Federal de Minas Gerais (UFMG) – Belo Horizonte, Brazil

{gabrielpoliveira,mirella}@dcc.ufmg.br

Abstract. *Music is one of the world's most important cultural forms, and also one of the most dynamic. Such a dynamic nature can directly influence artists' careers and reflect their success. In this work, we combine social networks and data mining techniques to analyze musical success from a genre-oriented perspective. Our goal is to mine exceptional collaboration patterns in success-based genre networks where the success deviates from the average. We conduct our analyses for global and eight regional markets, and the results show that each market has specific patterns of genre connections in which success is above average. Hence, our findings serve as a first step in developing strategies to promote future song releases across the world.*

1. Introduction

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. This widely adopted strategy is a strong force driving music nowadays, as such connections usually help artists bridge the gap between styles and genres, overlapping new fan bases and consequently increasing their numbers [Bryan and Wang 2011, Silva et al. 2022]. In fact, the genre perspective is very important when analyzing the impact of collaborations on musical success, as such partnerships may take place not only through intra-genre collaborations but also through inter-genres, bringing an additional dimension to hit songs [Mondelli et al. 2018, Oliveira et al. 2020].

Such a phenomenon may help to explain the popularization of regional genres in the global scenario. One of the biggest examples is the emergence of *k-pop* as a global genre. In April 2019, the collaboration between the group BTS and the American *pop* singer Halsey in the song *Boy With Luv* reached #8 on Billboard Hot 100 Chart. This and other collaborations with Western artists increased BTS' popularity in the United States and paved the way for the South Korean group to achieve #1 on Hot 100 with the following solo singles *Dynamite* (2020) and *Butter* (2021), which stayed in that position for 10 weeks. The success achieved by BTS and other groups such as Blackpink also shed light on other new *k-pop* acts that are becoming widely popular worldwide, including Tomorrow x Together and NewJeans.

In a lower-level analysis, such collaborations contribute to the emergence of local genres in the national market. For example, *pisadinha* is a new Brazilian genre that has become very popular in recent years. Such a genre has its roots in Brazil's Northeast region and is often considered a variant of the more traditional *forró*. To maintain and amplify its popularity in the Brazilian mainstream scene, *pisadinha* artists released several collaborations with artists from other popular genres. In 2020, Os Barões da Pisadinha collaborated with Xand Avião (a popular *forró* singer) in the song *Basta Você Me Ligar*. The song reached a huge success and entered the list of most listened songs on Spotify in the Brazilian (Top 2) and global (Top 50) charts.

Indeed, the rise of collaborations in the music market highlights its dynamic and unpredictable nature. Given the diversity of the music ecosystem and the collaborations between artists from distinct genres, it is challenging to conduct even descriptive analyses in such a context. Thus, in this paper, we go further in the study of genre collaborations by using genre networks and collaboration profiles to mine exceptional genre collaboration patterns in songs that have been successful in recent years. Specifically, we aim to achieve this goal through the following research question: *"In collaborative hit songs (i.e., with more than one artist), are there connection patterns between genres that achieve above-normal success?"* Our methodology combines social networks and data mining, and the results reveal distinct exceptional subgroups for each market, evidencing the importance of considering the local component in success analyses. After discussing related work in Section 2, our contributions are summarized as follows:

- From an existing dataset with success-based genre collaboration networks, we model the exceptional pattern task as a subgroup discovery problem (Section 3);
- We apply a subgroup discovery algorithm on the networks to mine such exceptional collaboration patterns (Section 4);
- We analyze the results in three perspectives, considering temporal and regional aspects: global, English-speaking, and non-English speaking markets (Section 5).

2. Related Work

There are several studies on factors that lead to musical success, creating an emerging field within computer science called Hit Song Science (HSS). It involves acquiring and analyzing musical data from distinct sources to study the relation between the intrinsic features of songs and their success [Pachet 2012]. Reinforcing its multidisciplinary characteristic, studies in HSS combine machine learning and data mining techniques with musicology and psychology concepts to verify whether popular songs share similar feature patterns. It was first introduced in 2003 by Polyphonic HMI,[1] a company that developed a commercial tool to predict, on a scale from 1 to 10, the success of a song in the current market based on its chart position.

Such an achievement motivated researchers in the academic community to develop initial scientific studies regarding hit song prediction. For instance, Dhanaraj and Logan [2005] use acoustic and lyric-based features in a classification model to provide the first evidence that there is indeed a pattern connecting hit songs, as their model performs slightly better than random. Following such a study, early works in HSS used mostly song-related features to predict popularity with distinct approaches. However, Pachet and Roy [2008] point out that the features commonly used at the time might not be enough to reveal relevant information about musical success. The advance of machine learning algorithms and the discovery of possible new features helped to overcome such obstacles [Mayerl et al. 2023, Zhao et al. 2023]. In addition, features extracted from social platforms (e.g., Twitter, Facebook and Instagram) have also been considered in prediction models [Araujo et al. 2017, Cosimato et al. 2019, Tsiara and Tjortjis 2020].

[1] Polyphonic HMI, Hit Song Science: `http://bit.ly/polyphonic-hmi`

Understanding musical aspects can be genre-dependent, and this also reflects in the musical success. Genre is one of the most prominent high-level music descriptors, and is fundamental within the musical scenario by aggregating songs that share common characteristics. Therefore, different studies in HSS use genre information in their models. For example, Shin and Park [2018] consider genres to understand the life trajectory of songs in Gaon Chart,[2] one of the main Korean music rankings. Regarding prediction models, Ren and Kauffman [2017] aggregate genres in a musical construct vector to summarize the acoustic content of a song. Such vectors are used as features in a regression model to estimate the popularity duration of a track from a top-charts perspective.

Besides most studies in HSS being related to prediction tasks, other works focus on further aspects of musical success. For instance, music networks have been used in studies aiming to analyze regional differences [Mondelli et al. 2018, Oliveira et al. 2020] and user preferences [Pereira et al. 2018]. Furthermore, collaboration-aware studies have become promising in HSS, as they provide 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, Silva et al. 2022].

Also, data mining techniques are used in the musical context on recognition of style hierarchies [Iloga et al. 2018], file indexing by style [Rompré et al. 2017] and music recommendation [Siddiquee et al. 2016]. Furthermore, Jorge et al. [2019] use subgroup discovery to detect unusual patterns in a network of social interactions. Therefore, combining the study of collaboration from a genre perspective with the application of data mining techniques can reveal important information about how artists from different communities come together to produce a new hit. Such a task is not free of challenges, but necessary to reveal new knowledge from the social perspective, as discussed next.

3. Methodology

This section presents the methodology used in this the work. From an existing musical dataset (Section 3.1), we use collaboration networks of genres in which the connections present distinct profiles (Section 3.2). Finally, we apply a subgroup discovery algorithm to verify exceptional collaboration patterns between different musical genres (Section 3.3).

3.1. Data

The dataset used is the Music Genre Dataset (*MGD*)[3] [Oliveira et al. 2020], which gathers information from Spotify, the most popular music *streaming* service today. *MGD* contains data on songs, artists and music genres based on global weekly *rankings* (comprises all markets where Spotify operates) and eight of the top ten music markets in 2019:[4] United States, Japan, United Kingdom, Germany, France, Canada, Australia and Brazil. Hence, this information allows to compare the world scenario and individual regional markets, in order to understand the relevance and dynamics of each country for music consumption and collaborations between artists.

For each market, there are weekly rankings of the 200 most streamed songs. The collection period is three years, covering every week between January 2017 and December 2019. In addition to these rankings, *MGD* presents information for each song (e.g.,

[2]On July 7, 2022, Gaon Chart was rebranded as Circle Chart.
[3]*MGD* is available for download at: `https://doi.org/10.5281/zenodo.4778562`
[4]IFPI Global Music Report 2019: `https://gmr.ifpi.org/`

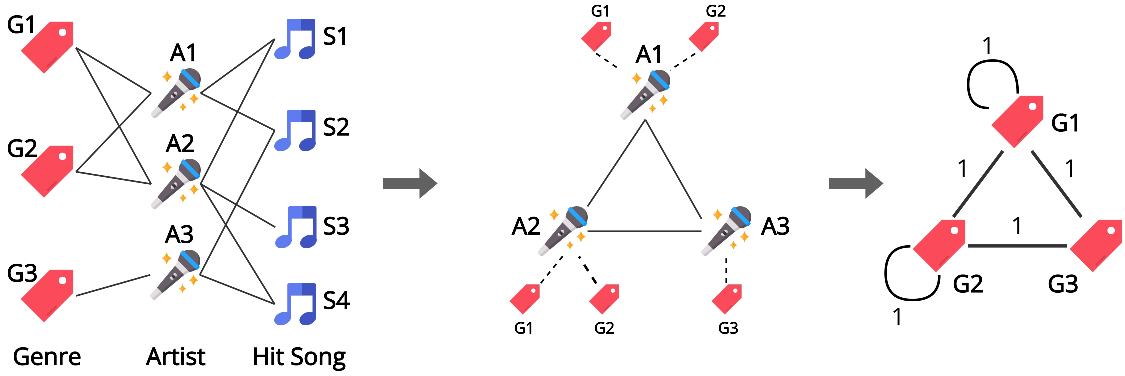

Figure 1. Building process of the Genre Collaboration Network.

release date, full list of performers and acoustic attributes) and artist. Overall, the dataset contains data from 1,370 rankings over 156 weeks, including 13,880 songs and 3,612 artists from 896 different music genres.

3.2. Genre Networks and Collaboration Profiles

To find exceptional behaviors in the connections between musical genres, we use the genre collaboration networks available in *MGD*. The dataset has individual networks for each market and year. The building process of the genre networks is illustrated in Figure 1. It starts from a tripartite graph, in which nodes represent songs, artists and genres. Note that, in Spotify, genres are linked to artists instead of songs. Then, the tripartite graph is reduced to an artist network (also available in *MGD*), which links artists who have collaborated in at least one hit song.

Finally, the artist networks generate the final genre networks. In such networks, nodes are musical genres, and edges connect the genres of artists who have collaborated on the same hit song. Edges are undirected and weighted by the number of songs involving artists from both genres (nodes/vertices). It is important to emphasize that *self-loop* edges can exist, since there are songs that are collaborations between artists of the same genre. For example, the song *Lençol Dobrado*[5] by João Gustavo & Murilo and Analaga generates an edge between them in the artist network; and each of the duo's genres (*sertanejo, electro* and *brazilian funk*) is linked to Analaga's only genre (*sertanejo*) with weight 1.

For each genre connection (i.e., network edge), we also consider its collaboration profile according to [Oliveira et al. 2020]. Such profiles allow to assess musical success by describing similar behaviors within collaborative songs from multiple angles. There are four distinct genre collaboration profiles in the networks: *Solid, Regular, Bridge,* and *Emerging*; each defined as follows.

Solid. It comprises partnerships established decades ago between the most popular super-genres. Examples include the collaborations between *pop* and *brazilian funk* artists, which are present in several hit songs which reached the top of the Brazilian charts.

[5]Most listened song on Spotify Brazil in 2019: `https://tecnoblog.net/noticias/2019/12/03/spotify-revela-musicas-artistas-mais-ouvidos-2019-decada/`

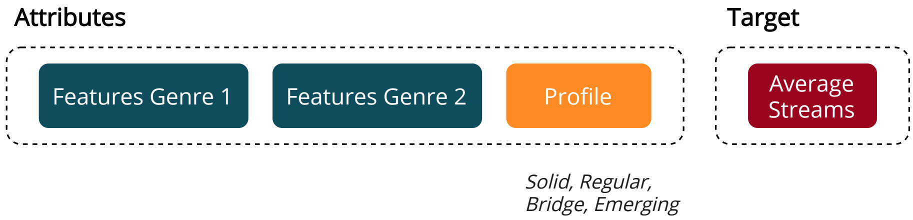

Figure 2. Representing the edges of Genre Collaboration Networks as instances of the Subgroup Discovery (SD) problem.

Regular. It contains the most common collaborations in all markets. They are similar to the *solid* ones, but they do not achieve the same success levels. An example of regular collaboration is the partnerships between *pop* and *forró*, which has become very popular in recent years, but is not as consolidated when compared to the previous profile.

Bridge. Such collaborations connect two distinct regions of the networks, and they represent possible investment targets, since they usually comprise genres with distinct audiences. In Brazil, the links of *gospel* with *rap* and *MPB* (Brazilian Popular Music) are good examples of bridge collaborations, since such genres do not have overlapping fan bases. Thus, a song with artists from these genres may achieve a high amount of streams.

Emerging. This profile contains mainly collaborations between regional genres and self-loop edges, i.e., collaborations intra-genre. They are called emerging because they usually occur between a well-established and an unknown artists. Examples of collaborations of regional genres here include *k-pop – k-pop* and *forró – forró*.

3.3. Network Subgroup Discovery

In order to properly answer our research question (i.e., to find genre connections with above-normal success), we use the genre collaboration networks from the previous section as the input of a Subgroup Discovery (SD) task. SD aims to identify relevant patterns (subgroups) that deviate from the standard. Its input is a dataset composed of a set of attributes describing the instances and a target variable, used to distinguish the behavior of the subgroups from the whole dataset. Thus, a subgroup is said to be exceptional if the distribution of the target in its instances deviates from the dataset [Klösgen and Zytkow 2002].

Subgroups are specified by a description language defined by domain experts and analysts. According to Rebelo de Sá et al. [2018], such languages are frequently composed of conjunctions of attribute conditions. For example, consider a dataset (different from *MGD*) in which the instances represent the music preferences of streaming users. For each person, the attribute set comprises demographic information including age, country, and marital status. Besides, the target variable is defined as the user's favorite music genre (e.g., pop, rock, or sertanejo). Thus, in a scenario where the overall favorite genre is *pop*, a possible subgroup found by an SD algorithm is:

$$\text{Age} \geq 40 \quad \wedge \quad \text{Country} = \text{“Brazil”} \quad \Rightarrow \quad \text{Genre} = \text{“Sertanejo”}$$

This subgroup means that the people over 40 who live in Brazil have a distinct genre preference when compared to the whole dataset. That is, Brazilians over 40 prefer

Table 1. Main acoustic features obtained from Spotify.

Feature	Description	Type	Value Range
acousticness	The probability of a song to be acoustic or not	Float	$[0, 1]$
danceability	Informs whether a song is suitable for dancing or not in terms of probability	Float	$[0, 1]$
duration_ms	The duration of a song in milliseconds	Integer	$[0, \inf)$
energy	The intensity and activity of a song considering information such as dynamic range, perceived loudness, timbre, onset rate, and general entropy	Float	$[0, 1]$
liveness	The probability of a song being performed live, i.e., the presence of an audience in a song	Float	$[0, 1]$
loudness	The general loudness measured in decibels (dB)	Float	Typically $[-60, 0]$
speechiness	The probability of a given song to have spoken words in it	Float	$[0, 1]$
tempo	The speed of the song, measured in beats per minute (BPM)	Float	N/A
valence	The positiveness of a song, in which high valence values represent happier songs, whereas low values means the opposite	Float	$[0, 1]$

sertanejo songs, whereas people in general are more into pop. Therefore, using SD in descriptive analyses helps to reveal hidden groups with exceptional preferences that deviate from the average.

In our approach (Figure 2), for each market and year,[6] we consider the network edges as the instances of the SD model, representing the collaboration between musical genres. Therefore, it is necessary to select attributes that describe the nodes individually and also characteristics already known from the collaboration. Hence, the attribute set for each instance is composed of features from the two genres of the respective edge, as well as the collaboration profile (*Solid*, *Regular*, *Bridge* and *Emerging*) for such an edge. To describe each genre, we select the following acoustic features from Spotify: *acousticness*, *danceability*, *duration_ms*, *energy*, *liveness*, *loudness*, *speechiness*, *tempo* and *valence*.[7] The definition of all features are presented in Table 1.

As such features are provided for individual songs, we assign to each genre, the median values of all songs from artists belonging to such a genre. These values are then discretized based on the quartiles for each variable. That is, values in the first quartile (below the 25th percentile) are classified as *low*, whereas values in the second and third quartiles (between the 25th and 75th percentile) are *medium*. Values above the 75th percentile are then classified as *high*. Finally, we set the average number of streams of each edge as our target variable, as it is a success metric provided by Spotify.

4. Experimental Setup

Here, we perform a subgroup discovery (SD) analysis in genre collaboration networks. We maintain the notion of temporality as we consider, for each market, a collaboration network from hit songs of each year (2017, 2018 and 2019). Hence, we analyze 27 distinct collaboration networks (three annual networks for nine music markets). Following the

[6] We consider each market and year separately to preserve temporal and regional aspects in our analyses.

[7] For full feature definition, see `https://developer.spotify.com/documentation/web-api/reference/#/operations/get-audio-features`

Table 2. Exceptional subgroups in the Global market (2017-19).

Market	Year	Subgroup	N	E	S_{SG}	S_D	Q
Global	2017	no exceptional subgroups found					
	2018	no exceptional subgroups found					
	2019	(G1: low acousticness, medium danceability, high degree, medium duration_ms, high energy) $\wedge$ (G2: medium energy, medium loudness, medium tempo) $\wedge$ regular profile	15	20	185.95	45.14	70.79

N: number of nodes **E:** number of edges $\mathbf{S_{SG}}$: average *streams* in subgroup (10^6)
$\mathbf{S_D}$: average *streams* in network (10^6) **Q**: quality metric value (10^7)

methodology presented in Section 3.3, we consider the network edges as the instances of our SD problem, described by acoustic features of the genres and the collaboration profile between them. We also set the average number of streams as the target value.

We use the *Beam Search* algorithm from the *pysubgroup* Python library[8] [Lemmerich and Becker 2018]. This algorithm finds the relevant subgroups according to a predefined target variable (here, the average number of *streams*), evaluated by a quality metric. In this work, we use the function *StandardQFNumeric* from the same library, which handles numeric target variables. For a given subgroup SG and a parameter α, this function is defined by Equation 1; where N_{SG} is the number of instances in the subgroup, N is the total number of instances in the dataset, μ_{SG} is the average of the target variable within the subgroup, and μ is the average of the target variable in the dataset. In our experiments, we empirically choose $\alpha = 0.5$, as we want to emphasize the difference in the target variable rather than the subgroup size.

$$q(SG, \alpha) = \left(\frac{N_{SG}}{N}\right)^{\alpha} (\mu_{SG} - \mu) \qquad (1)$$

5. Results

In this section, we answer our research question by finding exceptional genre collaboration patterns from hit songs, i.e., collaborations in which the success is above the average in the whole dataset. As language is crucial for listening to music, besides the Global market (Section 5.1), 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 (Section 5.2); whereas the latter comprises Brazil, France, Germany and Japan (Section 5.3). We finish this section by presenting an overall discussion on the results and the implications of this work (Section 5.4).

5.1. Global market

Table 2 shows the exceptional subgroups found for the global market. As the instances of our subgroup discovery problem are the network edges, we represent the attributes of each node (i.e., music genre) with the prefixes *G1* for the first genre and *G2* for the second. There are no exceptional subgroups in 2017 and 2018, which may indicate a

[8]pysubgroup: `https://github.com/flemmerich/pysubgroup`

homogeneous behavior in the collaborations during this period. Following the popularization of streaming services, there is a change in 2019, when a specific subgroup with an average stream count four times higher than expected emerges in the network. Such a subgroup includes edges in which the first genre has low acousticness, medium danceability, high degree (i.e., connectivity), medium duration, and high energy. In contrast, the second one has medium values for energy, loudness and tempo. In addition, the edges composing the subgroup share the *Regular* collaboration profile.

An example of genre connection within this subgroup is the collaboration between *electro house* and *pop*, which happens in the song *Happier* by Marshmello and Bastille. The song was released in August 2018, but its popularity grew until 2019, as it spent 27 weeks in the top 10 of the Billboard Hot 100. Furthermore, the collaboration between *electro house* and *synthpop* artists is also part of this exceptional subgroup. A good representative of song involving artists from both genres is *Here With Me* by Marshmello and the Scottish band CHVRCHES. This song charted at #31 on the Billboard Hot 100 and reached a peak of #11 on the Spotify Global Daily Chart, becoming one of the most successful singles of the band. Indeed, as of March 2023, it is still CHVRCHES' most popular song on Spotify, with over 600 million streams.

5.2. English-speaking markets

The results of *BeamSearch* for the networks from the English-speaking markets are presented in Table 3. In a first analysis, all regional markets have subgroups with attributes and target distributions different from each other. However, when analyzing each subgroup individually, we note some connections that repeat in some countries. For instance, in 2018, the edge between *dubstep* and *pop* belongs to exceptional subgroups in all such markets, which is reasonable since they share several cultural aspects. Besides, the Internet and social media advance provides a global platform where users can share and promote their musical tastes. The American DJ Marshmello appears again in two of the most popular songs involving *dubstep* (one of his musical genres) and *pop*: *Wolves* with Selena Gomez and *Friends*, in partnership with Anne-Marie.

In 2019, the exceptional subgroups in these markets are very similar to those in the global market. This similarity can be related to the significant participation of markets such as the United States and European countries in the global aggregate numbers. In other words, as the global charts are calculated from the absolute value of streams worldwide, the countries with the most listeners will have a greater weight in the final result. Thus, the subgroups found in these countries' networks also contain connections involving genres such as *dubstep* and *electro house*. Such genres had their success peak in the mid-2010s, and have seen a decline in their mainstream popularity in recent years. Overall, collaborations between such genres that reach a higher success are classified as exceptional, since they are no longer among the most popular genres today.

5.3. Non-English speaking markets

Table 4 presents the results for the four non-English speaking countries. In these markets, we highlight a strong presence of local genres in exceptional subgroups, revealing the popularity of such genres in their countries. For instance, popular subgroups in Japan involves connections of the *anime* genre with itself (i.e., intra-genre collaboration) and *j-pixie* in 2018 and 2019, respectively. In France, there is a strong presence of *reggae*

Table 3. Exceptional subgroups in English-speaking markets (2017-19).

Market	Year	Subgroup	N	E	$\mathbf{S_{SG}}$	$\mathbf{S_D}$	Q
Australia	2017	(G1: high danceability, low liveness) ∧ (G2: low liveness, high valence)	5	9	19.55	3.11	3.59
	2018	(G1: low acousticness, medium danceability, low valence) ∧ (G2: medium speechiness)	12	11	20.39	3.87	5.37
	2019	(G1: low acousticness, medium danceability, high degree) ∧ (G2: medium energy, medium loudness, medium speechiness) ∧ regular profile	14	19	11.05	2.41	3.97
Canada	2017	(G1: low liveness) ∧ (G2: medium acousticness, high danceability, medium degree, medium speechiness)	6	8	14.26	2.17	3.18
	2018	(G1: medium danceability, high energy, low valence) ∧ (G2: medium speechiness)	12	11	20.36	2.36	4.78
	2019	(G1: low acousticness, medium danceability, high degree, medium speechiness) ∧ (G2: medium danceability, medium loudness, medium tempo) ∧ regular profile	12	16	10.95	1.79	3.33
UK	2017	(G1: low liveness) ∧ (G2: medium acousticness, medium degree, medium duration_ms, low tempo)	11	17	27.23	4.19	7.73
	2018	(G1: low acousticness, high energy, low valence) ∧ (G2: medium speechiness, medium tempo)	11	10	37.80	6.61	8.72
	2019	(G1: low acousticness, medium danceability, high degree, medium liveness) ∧ (G2: medium danceability, medium loudness, medium tempo) ∧ regular profile	13	19	15.78	3.44	5.46
USA	2017	(G1: medium loudness, low tempo) ∧ (G2: medium duration_ms, medium speechiness, low tempo)	6	4	94.58	13.38	17.58
	2018	(G1: low acousticness, medium degree, high energy, low valence) ∧ (G2: medium speechiness)	13	12	85.66	14.20	22.03
	2019	(G1: low acousticness, medium danceability, high degree) ∧ (G2: medium energy, medium loudness, medium speechiness) ∧ regular profile	13	18	58.60	6.65	14.67

N: number of nodes **E:** number of edges $\mathbf{S_{SG}}$: average *streams* in subgroup (10^6)
$\mathbf{S_D}$: average *streams* in network (10^6) **Q**: quality metric value (10^7)

and *reggaeton* in the exceptional subgroups for 2019. Besides not being originated in France, both genres are becoming widely popular in that country. Such a vibrant scene and influences of artists with roots in the North of Africa are paving the way for the emergence of new genres, such as *francoton*.

Furthermore, Brazilian regional genres appear in subgroups from all considered years. Specifically, in 2019, the subgroup that comprises the edge between *afrofuturism* and *pagode baiano* has an average number of streams more than ten times bigger than the whole network. Such a huge success is boosted by songs such as *Bola Rebola* by Anitta, Tropkillaz, J Balvin and MC Zaac, which debuted in #1 in Brazil's daily chart in Spotify[9] with more than 1.2 million streams. Therefore, as the second and tenth biggest music markets in the world,[10] such countries reinforce the importance of considering regional

[9]Spotify Daily Chart, 2019/02/22: `https://charts.spotify.com/charts/view/regional-br-daily/2019-02-22`

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

Table 4. Exceptional subgroups in non-English-speaking markets (2017-19).

Market	Year	Subgroup	N	E	S_{SG}	S_D	Q
Brazil	2017	(G1: medium danceability) ∧ (G2: high acousticness, medium degree, low duration_ms, high energy)	14	19	18.69	6.13	5.14
	2018	(G1: high speechiness) ∧ (G2: high acousticness, high danceability, low duration_ms)	3	2	96.95	4.82	12.79
	2019	(G1: medium acousticness, high liveness, high speechiness) ∧ (G2: high danceability, high valence)	4	3	67.08	4.41	8.33
France	2017	(G1: medium acousticness, medium degree, low tempo) ∧ (G2: medium duration_ms)	12	25	8.28	2.74	2.45
	2018	(G1: medium danceability, medium duration_ms, high liveness, high loudness) ∧ (G2: medium tempo)	14	19	8.50	2.91	2.54
	2019	(G1: high danceability, medium liveness, high valence) ∧ (G2: medium acousticness)	3	2	27.93	4.16	3.24
Germany	2017	(G1: low liveness) ∧ (G2: medium acousticness, medium degree, high energy, medium speechiness)	6	8	35.92	3.65	7.79
	2018	(G1: low acousticness, high loudness, high tempo, low valence) ∧ (G2: medium tempo)	12	11	23.84	4.45	6.87
	2019	(G1: high energy) ∧ (G2: low liveness, low tempo)	9	22	20.03	3.87	5.36
Japan	2017	(G1: medium danceability, medium liveness) ∧ (G2: medium danceability, medium duration_ms, medium loudness, low tempo) ∧ regular profile	11	16	0.76	0.27	0.19
	2018	(G1: high loudness) ∧ (G2: low danceability) ∧ solid profile	2	2	8.36	0.35	1.11
	2019	(G1: low acousticness, low danceability, low speechiness) ∧ (G2: low valence)	3	2	12.30	0.42	1.62

N: number of nodes **E:** number of edges $\mathbf{S_{SG}}$: average *streams* in subgroup (10^6)
$\mathbf{S_D}$: average *streams* in network (10^6) **Q**: quality metric value (10^7)

markets individually, as their engagement shapes the global environment.

5.4. Discussion

The use of subgroup discovery (SD) in music genre networks is an important tool to reveal collaborations between genres that are not very obvious, but an excellent investment opportunity due to their above-average success. Such an opportunity can be explained from the very definition of SD, since exceptional subgroups will not reveal collaborations that are known to be successful (e.g., pop and rap). Therefore, subgroups are valuable information for artists and record labels to explore new partnership opportunities and innovate within the music industry. Furthermore, emerging music genres can be better promoted to achieve the mainstream, and artists have the chance to reach new audiences.

Overall, our results corroborate previous works that say that collaboration in the music scene can influence musical success. By adding the musical genre perspective to the collaboration study, we contribute to increasing the knowledge about factors behind musical success and the phenomenon of collaborations in the music industry, which is only possible due to social network analyses. In addition, our findings reinforce the importance of analyzing success data from regional markets since each market has different patterns of success compared to each other and the global market.

6. Conclusion

In this work, we combine social network analysis and data mining techniques to uncover exceptional patterns in musical genre collaborations. We use a Subgroup Discovery technique in genre networks to detect genre connections in which the success metric (i.e., the average number of streams) deviates from the whole network. Our results show exceptional subgroups in all markets, and each one presented distinct results that show the importance of considering the local component in success analyses. Following findings from previous works, regional markets behave differently compared to the global scenario or even to the United States, which is the biggest music market in the world.

Therefore, we can answer our research question by revealing that there are indeed genre connection patterns in hit songs that achieve above-normal success. Our findings provide benefits to both artists and record labels, as they can diversify their partnerships and plan future releases. Further, we emphasize the use of Social Network Analysis onto a distinct, valuable industry – music. For example, artists from regional genres such as *k-pop* and *pisadinha* may choose to collaborate with the most promising genres to achieve a higher level of success, achieving a breakthrough status and reaching broader audiences.

Limitations and Future Work. We plan to evaluate other subgroup discovery algorithms and quality metrics, in addition to expanding the regional and temporal coverage of the dataset. In addition, future work should investigate in depth the impact of genre collaboration on individual success. Furthermore, we plan to consider other features to describe music genres in our SD model.

Acknowledgments. This work was funded by CAPES, CNPq and FAPEMIG.

References

Araujo, C. V. et al. (2017). Predicting music success based on users' comments on online social networks. In *WebMedia*, pages 149–156, Brazil.

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

Cosimato, A., Prisco, R. D., Guarino, A., Malandrino, D., Lettieri, N., Sorrentino, G., and Zaccagnino, R. (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.

Iloga, S., Romain, O., and Tchuente, M. (2018). A sequential pattern mining approach to design taxonomies for hierarchical music genre recognition. *Pattern Anal. Appl.*, 21(2):363–380.

Jorge, C. C., Atzmueller, M., Heravi, B. M., Gibson, J. L., Rebelo de Sá, C., and Rossetti, R. J. F. (2019). Mining exceptional social behaviour. In *EPIA (2)*, volume 11805 of *Lecture Notes in Computer Science*, pages 460–472. Springer.

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

Lemmerich, F. and Becker, M. (2018). pysubgroup: Easy-to-use subgroup discovery in python. In *ECML/PKDD (3)*, volume 11053 of *LNCS*, pages 658–662. Springer.

Mayerl, M., Vötter, M., Specht, G., and Zangerle, E. (2023). Pairwise learning to rank for hit song prediction. In *BTW*, volume P-331 of *LNI*, pages 555–565.

Mondelli, M. L. B., Gadelha Jr., L. M. R., and Ziviani, A. (2018). O que os países escutam: Analisando a rede de gêneros musicais ao redor do mundo. In *BraSNAM*, Natal.

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.

Pachet, F. (2012). Hit song science. In Tao Li, Mitsunori Ogihara, G. T., editor, *Music Data Mining*, chapter 10, pages 305–326. CRC Press, New York, NY, USA.

Pachet, F. and Roy, P. (2008). Hit song science is not yet a science. In *ISMIR*, pages 355–360, Philadelphia, USA.

Pereira, F. S. F., Linhares, C. D. G., Ponciano, J. R., Gama, J., de Amo, S., and Oliveira, G. M. B. (2018). That's my jam! uma análise temporal sobre a evolução das preferências dos usuários em uma rede social de músicas. In *BraSNAM*, Natal. SBC.

Rebelo de Sá, C., Duivesteijn, W., Azevedo, P. J., Jorge, A. M., Soares, C., and Knobbe, A. J. (2018). Discovering a taste for the unusual: exceptional models for preference mining. *Mach. Learn.*, 107(11):1775–1807.

Ren, J. and Kauffman, R. J. (2017). Understanding music track popularity in a social network. In *25th European Conference on Information Systems*, pages 374–388, Atlanta, GA, USA. AIS.

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

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

Siddiquee, M. M. R., Rahman, M. S., Chowdhury, S. U. I., and Rahman, R. M. (2016). Association rule mining and audio signal processing for music discovery and recommendation. *Int. J. Softw. Innov.*, 4(2):71–87.

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, Curitiba, Brazil.

Tsiara, E. and Tjortjis, C. (2020). Using twitter to predict chart position for songs. In *IFIP Artificial Intelligence Applications and Innovations*, pages 62–72, Neos Marmaras, Greece.

Zhao, M., Harvey, M., Cameron, D., Hopfgartner, F., and Gillet, V. J. (2023). An analysis of classification approaches for hit song prediction using engineered metadata features with lyrics and audio features. In *iConference (1)*, volume 13971 of *Lecture Notes in Computer Science*, pages 303–311.