From Exploration to Exploitation: Understanding the Evolution of Music Careers through a Data-driven Approach

Danilo B. Seufitelli, Mirella M. Moro

Department of Computer Science
Universidade Federal de Minas Gerais (UFMG) – Belo Horizonte, MG – Brazil

{daniloboechat,mirella}@dcc.ufmg.br

Abstract. *We propose a data-driven methodology to analyze how music artists spread their interest (explore) and focus their attention (exploit) on distinct music topics (genres) while having peaks of success (hot streaks). Music topics are identified through community detection over a complex network modeling of artists, songs and their genres. Then, we analyze exploration and exploitation by measuring the entropy and quantifying it in three periods: before, during and after hot streaks. Results show artists explore topics before hitting their first hot streak and during their most successful period. After, they tend to reduce the range of topics by going deeper in a handful only. Such findings are relevant for identifying and nurturing creative talent in the music industry.*

1. Introduction

The music industry has benefited from a long, productive collaboration with computer science, and specially Databases. Indeed, many data-oriented techniques provide proper interface for accessing large datasets on music and artists, which, in turn, has boosted research in Music Information Retrieval (MIR). Since Byrd and Crawford's (2002) position paper, many music industry problems use data-oriented solutions to classifying music genres, detecting collaboration profiles and so on (e.g., [Oliveira et al. 2020]).

Despite such advances, the music industry still requires specialized approaches due to data-oriented challenges, such as dynamic evolution, volatile information, and diverse data sources, which instigate music managers to consider the power of data and database tools. A meaningful open issue is music success: *Why do careers become successful? How to build successful careers? How to maintain success?* To solve such questions, researchers may use database techniques such as those based on time series. Going one step forward (or above), studying creative careers (e.g., music, literature) offers many interesting and relevant questions that can be explored over their data perspective.

Creative careers are interconnected with several factors that impact their success, including luck, hard work and resilience [Yucesoy and Barabási 2016, Passman 2019]. Still, there is no magic formula or off-the-shelf recipe for success in creative careers. Indeed, a successful career may have to undergo several phases and stages until it consolidates itself in the market. Such characteristic is also present in science, film, music, art, and even sports [Yucesoy and Barabási 2016]. Aside from "regular" success, identifying the most impacting work or the most successful period within a creative career is challenging. These most successful periods are known as *Hot Streaks*: a specific period during which an individual's performance is substantially better than their typical performance [Liu et al. 2018]. Further, Seufitelli et al. (2022) have already shown that it is possible to identify the presence of success above the normal in the musical context.

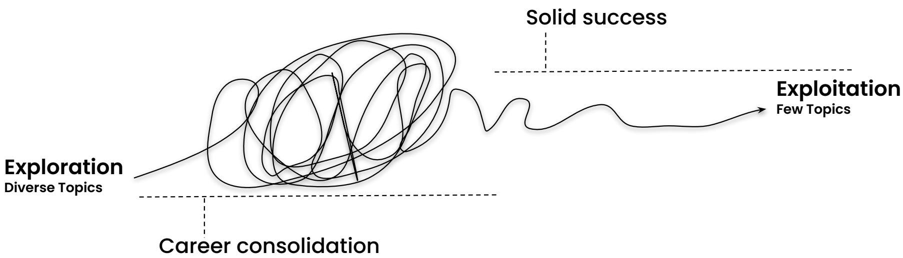

Figure 1. Exploration and Exploitation Phases.

Regarding data, the music industry is a unique source of complex perspectives. First, the industry dynamics is always evolving – for example, a huge upgrade has changed it from the physical era (lead by physical LPs, CDs, and DVDs) to the current digital age (lead by streaming platforms). Second, the phonographic market is complex and formed not only by musicians but also by record companies, labels, studios, engineers, and producers. Further, another challenging dimension considers the type or genre of music, which divides the whole market into specific niches, from classic jazz to k-pop.

The latter perspective is key, as artists in the early stages of their careers usually explore different topics and genres. They explore features that allow them to transit among different target audiences. For example, young artists may explore new rhythms before identifying with a specific one. As their careers take off, artists tend to specialize more in a particular music segment, with less market exploration. This behavior leads us to question whether the above-normal success links to two concepts: Exploration and Exploitation.

In summary, *Exploration* refers to the process of visiting new regions of the search space; whereas *Exploitation* is the process of delving into a given area in search of a local optimum. Another important concept is *entropy*, which measures the diversity or variability within a dataset and helps to evaluate the balance between exploration and exploitation. These concepts are widely discussed in the field of artificial intelligence and are known to be the pillars of solving search problems [Črepinšek et al. 2013]. Then, bringing such concepts to music careers, we hypothesize that artistic careers follow the exploratory phase until they reach their peak of success (hot streak). Specifically, Figure 1 shows a representation of the exploration and specialization phases: at the beginning of their careers, artists may explore different topics when they are still consolidating themselves in the market (different rhythms, genres, audiences, presentation formats, etc); as their careers are consolidated and known, they tend to focus on whatever works better.

Putting all together, it is still unclear if and how hot streaks relate to topics/genres explored and exploited by music artists. In other words, to the best of our knowledge, we are the first to investigate topics *entropy* as a possible factor for Hot Streaks. Overall, our work is motivated by two key questions:

RQ1: *What are the existing topics (genres) in a given musical career?*

RQ2: *Do musical careers reflect exploitation or exploration regarding such topics within a timeline of hot streaks?*

Our contributions are summarized as follows. First, we describe a general methodology that provides a clear strategy for analyzing exploitation and exploration over hot streaks timeline (Section 3). Next, as the music topics are the core of entropy analysis, we propose and build a song network to extract the work topics of each artist (Section 4). The goal is to answer RQ1 by extracting relevant topics that minimally explain musical careers and measuring their diversity through time (before, during, and after the careers achieve hot streaks). Next, we answer RQ2 by analyzing entropy to understand musical career exploration and exploitation dynamics (Section 5). Such research may help understand how artists can manage their careers at different times, with the peak period of above-average success (i.e., their first hot streak) as the point of observation. Finally, we go over conclusions, future work and limitations (Section 6).

2. Related Work

Music is an important form of cultural expression and has one of the most dynamic industries. Over recent decades, technology advances have changed how people consume music, from physical records to streaming services. Hence, music has a distinct social dimension, whether shaped by the engagement of artists or listeners. Distinct social media services (e.g., Facebook, YouTube, and Twitter) are designed to engage audiences and encourage them to discover new artists, share recommendations and consume music [Amorim et al. 2022]. Accordingly, the number of studies aimed at discovering the recipe for musical success has increased, defining the field of *Hit Song Science* (HSS), which focuses on predicting the popularity of songs [Pachet 2011]. Then, recent studies analyze the impact of different factors on musical success, including acoustic characteristics [Martín-Gutiérrez et al. 2020], collaboration [Silva et al. 2019, Oliveira et al. 2020], or even engagement on social platforms [Cosimato et al. 2019].

Creative careers have always instigated researchers and industry to understand how to build or improve them. Recent studies use machine learning techniques to propose algorithms, approaches and studies to include or exclude features that enhance the prediction task, e.g., [Baldo et al. 2022]. Consequently, the music industry benefits from the researchers' proactivity, as they always propose updates regarding the phonograph market nuances to computational algorithms. However, we note a lack of exploring the common characteristics that lead to successful musical careers and associating such attributes with the most relevant periods in the artists' careers. After all, a significant milestone in artistic careers is the existence of Hot Streaks, that is, those periods of high notoriety and presence in the music market. Although Hot Streaks are practically a rule in musical careers, it is unclear if there are patterns regarding their beginning. The lack of systematic explanations for Hot Streaks and the randomness of when they occur within music careers support an unpredictable view of such careers.

Recently, [Liu et al. 2021] have explored the characteristics that contribute to the emergence of hot streaks in three different creative careers: scientific, cultural, and artistic. They find Hot Streaks are not associated with either exploring or exploiting behavior in isolation. Furthermore, they show real careers are complex, with heterogeneous influences operating across domains and many individual and institutional factors. In this context, we understand musical careers are similar to such creative careers and lack studies to characterize better what precedes the periods of Hot Streaks. Therefore, our goal is to understand how music careers achieve success by exploring and exploiting (entropy) topics in their careers. Such knowledge complements the results of machine learning tasks that create a model for predicting success. Further, we generate such information by considering technical and social characteristics as well.

Figure 2. *The Weeknd* success time series over the Global market.

3. Methodology

Understanding what leads an artist to reach their most successful period (or the period in which their career gets notoriety) requires evaluating data-driven perspectives of the music domain that evolve over time. One major angle is the musical genre of artists and their songs. Hence, we introduce a methodology to extract topics from artists' careers from the genres associated with them. In order words, we seek the topics (in this case, genres) associated with each song released by each artist. However, we understand that music genre classification is a complex problem in the Music Information Retrieval area, and is beyond our scope. Overall, the new methodology has five steps described ahead.

Collect Data. We built a *crawler* in Python using the Spotify API[1] to collect the data. First, we collected both Global and Brazil's Top 200 Daily chart data from Spotify Charts. For both cases, the collection date started on 2017-01-01 (Spotify's first availability date) and ended on 2022-03-13. After this period, Spotify closed access to its chart data. For both the Global and Brazilian markets, we got 379,200 chart lines. We then collected data for all artists, resulting in 286,275 artists for the Global market and 217,749 for Brazil. Finally, we collected data from all songs found on the charts, with 2,195,243 songs for the Global market and 1,447,784 for the Brazilian market.

Build Time Series. The evolution of an artist's success is represented by the number of daily streams received on Spotify. We use Global and Brazil's Spotify Top 200 Chart to model artists' success over time. For all artists, each point in their time series represents the total number of streams (i.e., the number of times the song was listened to on Spotify), which is our success measure. For example, Figure 2 presents the time series of *The Weeknd*. The album *After Hours*, released in 2020, continues to break records even after years. In addition to breaking all records with *Blinding Lights*, which appeared for 90 weeks on the most important chart in the United States (Billboard), the band also celebrates *Save Your Tears* with 60 weeks of charting on the Billboard Hot 100.[2]

Detect Hot Streak. From the time series, we can now detect the artists' most successful periods, i.e., their Hot Streaks. We do so by applying the PAA method – Peacewise Aggregate Approximation, which reduces the dimensionality of the time series in N new dimensions [Keogh and Pazzani 2000]. Given a time series $X = x_1, x_2, \cdots, x_n$ of length

[1]Spotify API: `https://developer.spotify.com/`
[2]The Weeknd Charts: `https://bit.ly/theWeeknd_history`

n, PAA reduces it into a new series $\overline{X} = \overline{x_1}, \overline{x_2}, \cdots, \overline{x_N}$ with N dimensions, $1 \leq N \leq n$. The intuition is that dividing the original time series into N segments of equal size produces N new points. The value of each segment is defined as the average of the points within that frame (Equation 1). Thus, each point of the original time series is approximated by assigning the PAA value of its respective segment. This method helps balance days with little or no success metric values, which was quite common in the data. Its only parameter is the number of segments to divide the series into. After empirical tests, we set this parameter to 90 days ($N = 90$): a value that captures enough information to represent an artist's success patterns over time, without being too short that could lead to information loss and scant representation of an artist's career trajectory. Next, we define a Hot Streak as the period when success (i.e., number of streams) is above a certain threshold obtained from the career. In other words, hot streak detection does not consider external factors (e.g., genre and time) because artists reach different levels of success, and choosing a single threshold would make the comparison unfair.

$$\overline{x_i} = \frac{n}{N} \sum_{j=\frac{n}{N}(i-1)+1}^{\frac{n}{N}i} x_j \qquad (1)$$

$$\tilde{H} = -\sum_{i=1}^{m} p_i \, log \, p_i \qquad (2)$$

Build Songs Network. With the hot streak information from artists' careers, the next step is building a music network to extract topics from their careers before, during, and after their Hot Streaks. Given two songs A and B as nodes, there is an edge between them if there is an interposition of genres between artists of A and B (detailed in Section 4). Spotify provides a set of genres for each artist; and such network gives information on musical genres associated with each artist. As we want to analyze the evolution of each career in terms of topics (genres) explored, we need one or more topic per song.

Analyze Entropy. Finally, the last step of the methodology is to apply an algorithm that calculates the degree of entropy of the distribution of topics in the artists' careers. The intuition is if an artist explores many topics, they will be in the exploration phase; otherwise, by focusing on a handful of topics, the artist is in the exploitation phase.

To quantify the exploration and exploration behaviors reflected in the careers of musical artists, we measure the style or entropy of the topic of their music (i.e., their artistic productions). We define it as the frequency an artist engages in an art style or topic, and is the number of unique styles or topics, according to Equation 2 – where i is the set of topics (genres) explored by an artist, p_i is the frequency of "devotion" by an artist, and m is the set of unique genres of whole artist's career. If an artist uses a pure exploitation strategy ($\tilde{H} = 0$), their work sits in only one style or topic (i.e., the genre tuple returned by the LDA in the next section). Otherwise, an artist is in pure exploration ($\tilde{H} = log\, n$) when they divide attention equally in more styles or topics.

4. Songs' Network Modeling

Studying exploration and exploitation of music careers requires extracting topics that artists work on or have worked on to create a timeline of the spectrum of explored topics. Still, our dataset[3] has no genre associated with each song/artist's work; it has an aggre-

[3]MUHSIC: `https://doi.org/10.5281/zenodo.5591015`

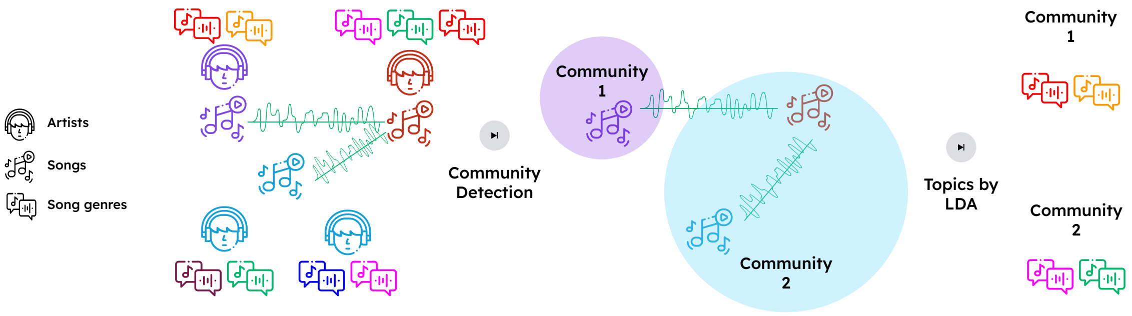

Figure 3. Topic Modeling to Song's artists.

Algorithm 1 Song network modeling algorithm

Input: Songs list (S) and artist genres list (G_L)
Output: Song network G

```
Create an empty edge list E
for each song s_i in S do
    Add s_i in g_i                          ▷ Assign the respective artist's genre set g_i ∈ G_L
end for
for each pair of songs s_a and s_b ∈ S do
    if len(g_a ∩ g_b) > 0:                  ▷ Check if the intersection between genres set g_a and g_b is not empty
        create s_a <-e_i-> s_b              ▷ Add and edge e_i between s_a and s_b with respective genres set g_a and g_b
end for
```

gated set of all genres already explored by artists, as Spotify provides a list of genres for each artist. Then we propose a method to classify musical genres with a song network.

4.1. Topic Modeling Methodology

Figure 3 presents the method for extracting artists' topics, as defined in Algorithm 1. First, it builds a music network based on the genres of all artists involved in the songs. Let S be a set of nodes represented by each song in our dataset, and E a set of edges connecting the songs. There is an edge e_i between two songs s_i and s_j if there is an intersection (partial or total) of the genre list of all participating artists of the respective songs s_i and s_j.

From the music network, the next step is applying a community detection algorithm to identify each community's "top" genres. In other words, all songs that are part of a given community will have the most common genre of the respective community. The proposed solution uses Louvain's community detection algorithm,[4] a simple method to extract the community structure of a network by using Python's NetworkX library.

Finally, the solution applies the Latent Dirichlet Allocation (LDA) algorithm to extract the most relevant genres from each recognized community. LDA is a generative probabilistic model of a corpus with the objective of modeling topics. It works on the premise that each topic is a set of terms, and each document is a mixture of a set of topics. Then, documents represent random mixtures over latent topics, where each topic characterizes a distribution over words [Blei et al. 2003]. We use two to model the communities' topics as the number of terms returned for each topic (i.e., a pair of song genres as a topic for each community). We also configure the algorithm to return only one topic per community. By selecting two terms per topic for each community, the result is a concise and interpretable representation of the genres associated with the community. Also, pairs allow a direct comparison between communities in terms of their musical core.

[4]Networkx Louvain's Community: `https://bit.ly/networkx_louvain`

Table 1. Music network metrics of Global Market.

Metric	Value
Number of nodes	7,725
Number of edges	15,492,869
Average degree	4011.09
Assortativity	-0.018
Average centrality degree	0.52
Density	0.52

Table 2. Music network metrics of Brazilian Market.

Metric	Value
Number of nodes	4,933
Number of edges	4,522,166
Average degree	1833.43
Assortativity	0.14
Average centrality degree	0.37
Density	0.37

4.2. Topic Modeling Results

We now present the results of this phase, with the network metrics for global and Brazilian markets in Tables 1 and 2, and LDA topics for both markets in Tables 3 and 4.

Global Market. We build a network for extracting the artist's topics from the Global market. As an example of subnetwork, consider the songs: *Stay* by The Kid LAROI ft. Justin Bieber; *God Is A Dancer* by Tiësto ft. Mabel; and *Look at Me!* by XXXTENTACION. The artist Justin Bieber has the genre *pop* on Spotify; The Kid Laroi has *hip hop*; Tiësto has *trance, house, dubstep, dance-pop, electro, pop, tropical house*; Mabel has *dance pop, pop, electro, tropical house*; and XXXTENTACION has *rap, hip hop*. The song *Stay* is assigned to the pair of genres [*pop, hip-hop*]; *God Is A Dancer* to [*tropical house, dubstep, trance, electro, house, dance-pop, pop*]; and song *Look at Me!* to [*hip hop, rap*]. Hence, after applying the network modeling, there is a link between the songs *Stay* and *God Is A Dancer* because the genres lists for both songs have pop in common. Likewise, there is a link between *Stay* and *Look at Me!* because of the hip-hop intersection. The main statistics of the network created are summarized in Table 1. As the network is built by pairwise combinations, it has 15,492,869 edges and 7,725 nodes.

The next step is to identify the communities in the generated network. The Louvain algorithm has the *resolution* parameter – a key factor in community detection as it controls the level of granularity at which communities are identified.[5] After empiric experiments, we set such a parameter to 1.15, and Louvain returned 16 communities. These communities represent unique groupings of nodes in the network that are tightly interconnected with one another, providing valuable insights about the main genres for the songs set in each community, delimited by the LDA algorithm.

Finally, finding the main topic of each community is key in defining a genre for all songs within that group. Therefore, we employ the LDA algorithm, which allows to isolate the most relevant themes present in each community. Overall, LDA returns a pair of genres representing each community's main topic; which then becomes the primary genre for all songs associated with that community. For instance, community zero's songs share the [k-pop, anime] genre, while community 12's songs fall under the [reggaeton, latin] genre. Table 3 shows the topics produced by the algorithm for each global market community. The songs used as examples were set as part of the following communities: *Stay* in community 13, with the topic [*pop, grime*]; *God is a Dancer* in community 1, with [*dance pop, electro*]; and *Look at Me!* in community 4, with [*hip hop, pop*].

[5]When the value of *resolution* is less than one, the algorithm favors identifying larger communities; whereas values greater than one may jeopardize the detection of smaller, more specific communities.

Table 3. LDA topics for Global Market communities.

ID	LDA Topic	ID	LDA Topic
0	[k-pop, anime]	8	[new wave, classic rock]
1	[dance pop, electro]	9	[arrocha, sertanejo]
2	[dance pop, hip hop]	10	[rap, pop]
3	[hip hop, rap]	11	[pop, trap]
4	[hip hop, pop]	12	[reggaeton, latin]
5	[r&b, rap]	13	[pop, grime]
6	[hip hop, trap]	14	[psychedelic]
7	[electro, trap]	15	[mariachi, ranchera]

Table 4. LDA topics for Brazilian Market communities.

ID	LDA Topic
0	[rock, mpb]
1	[k-pop, k-rap]
2	[pop, dance pop]
3	[brazilian funk, pop]
4	[sertanejo, arrocha]
5	[hip hop, rap]
6	[easy listening, lounge]

Brazilian Market. Similarly, consider the following songs as examples of a subnetwork from the Brazilian Market: *Mal Feito - Ao Vivo* of Marília Mendonça ft. Hugo & Guilherme; *Piseiro Estourou - Ao Vivo* of Os Barões Da Pisadinha; and *Avisa Que Eu Cheguei* of Naiara Azevedo ft. Ivete Sangalo. Artists Marília Mendonça and Hugo & Guilherme have the genres *sertanejo, arrocha* on Spotify; Os Barões da Pisadinha has *arrocha, forro*; Naiara Azevedo has *sertanejo, brazilian funk*; and Ivete Sangalo has *axe, samba reggae, pagode, arrocha, brazilian funk, mpb, pop*. Consequently, the song *Mal Feito* is assigned to the pair of genres [*arrocha, sertanejo*], the song *Piseiro Estourou* to [*forro, arrocha*], and the song *Avisa Que Eu Cheguei* to [*brazilian funk, sertanejo, samba reggae, mpb, pop, axe, pagode, arrocha*]. After applying the network modeling, there is a link between the songs *Mal Feito* and *Piseiro Estourou* because the genre lists for both songs have *arrocha* in common. Likewise, there is a link between *Piseiro Estourou* and *Avisa Que Eu Cheguei* because of the *arrocha* intersection. The main statistics of the network created are summarized in Table 2. Similar to the global market, as the network is built by pairwise combinations, it is expected to have many edges. Such a network has 4,522,166 edges, and the number of nodes is 4,933.

With the network created, we also run an empirical test to find the *resolution* parameter that controls the level of community granularity. We set this parameter to 1.3, resulting in seven distinct communities being identified. Hence, we also apply the LDA to obtain a pair of genres representing each community's main topic. We then designate this pair as the primary genre for all songs associated with that community. For instance, community zero's songs share [rock, mpb] genres, while community 2's songs are set in the [pop, dance pop] genres. Table 4 shows the topics produced by the algorithm for each global market community. Here, the songs *Mal Feito, Piseiro Estourou* and *Avisa Que Eu Cheguei* previously used as examples were set as part of community 4 [sertanejo, arrocha]. Once each community has its primary genres, the next phase of our methodology is to apply entropy analysis to explore the topics covered by the artists.

5. Entropy Analyses

Countless factors can influence an artist's career evolution and success. The exploration and exploitation strategies have attracted interest in a wide range of expertise areas, leading us to study their potential relationship to hot tracks in the musical ecosystem. First, we need to distinguish these two topics: exploration and exploitation. *Exploitation* allows individuals to build knowledge in a given area and refine their capabilities over time. Such a phenomenon may help understanding how to achieve hot streaks, as exploitation allows

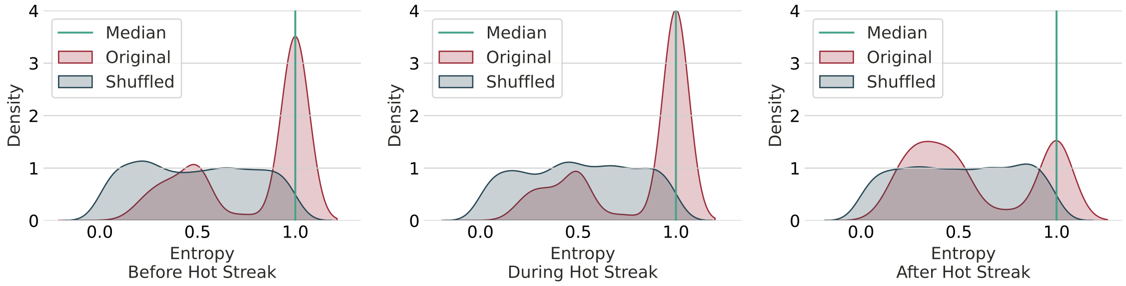

Figure 4. Entropy distribution of Global Market.

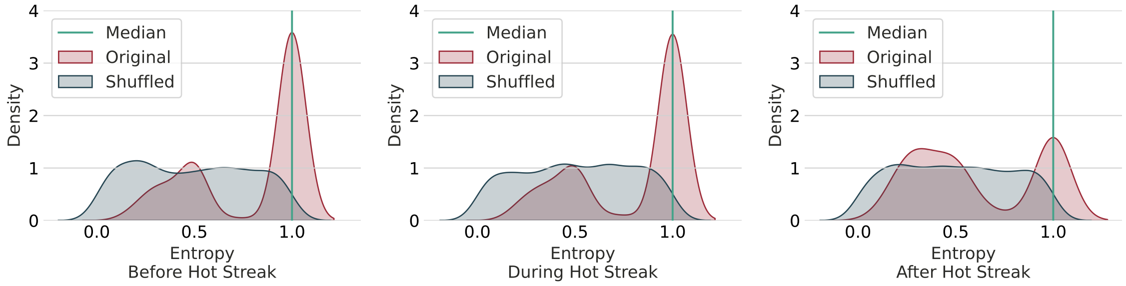

Figure 5. Entropy distribution of Brazilian Market.

individuals to focus on a particular "area" to establish expertise in and gain a reputation related to such proficiency. For musical artists, we may say that an artist focuses on a small range of musical genres, for example. On the other hand, *exploration* involves individuals experimenting and searching beyond their existing or previous competence areas. Defining a parallel with music, artists explore various musical genres in their careers.

Furthermore, whereas exploration is riskier and consequently associated with more significant variation in results, it can also increase the likelihood that someone will stumble upon an innovative idea through unforeseen combinations of disparate sources. In contrast, exploitation can suppress originality and limit an individual's ability to consistently produce high-impact work over time. In this sense, the benefits and disadvantages of these contrasting approaches raise an essential question: Are the behavior of musical artists' careers a reflection of exploitation or exploration? Understanding the balance between these two strategies and how they contribute to the success of artists can help acquiring insights into the nature of creativity and innovation in the musical industry.

Indeed, the results of our study reveal an interesting pattern in the career trajectories of successful musical artists. Specifically, music artists tend to diversify their musical styles before and during their first Hot Streak period. This trend is illustrated in Figures 4 and 5, which depict the entropy distribution in three stages of an artist's career: before, during, and after their Hot Streak, for both the Global Market and the Brazilian Market. Note we also plot and compare artists' careers with random careers to check the robustness of the results. In other words, random careers are generated by the computer to compare with real careers and verify the differences.

The graphics show that artists are more likely to explore fewer new rhythms and musical genres after achieving above-normal success (i.e., exploitation phase). Furthermore, our data revealed that some artists continue to invest in exploration even after their Hot Streak period has ended, indicating that they view exploration as a crucial aspect of their creative process. This behavior suggests that artists may feel more inclined to experiment with new sounds and styles when they have achieved a certain level of recognition and have a broader platform to showcase their work.

A possible scenario for artists who continue to explore new sounds is: they use their fame and influence to support emerging artists who work in different genres. Successful artists can continue trying and exploring new styles while supporting new talent by starring contemporary artists at the beginning of their careers. This approach also helps keeping their work fresh and relevant as they face new ideas and sounds through collaborations with emerging artists. Overall, these findings provide valuable insights into successful musical artists' strategies to sustain their success in the industry. By diversifying their musical styles and investing in exploration, these artists are better equipped to adapt to changing trends and maintain their relevance in a constantly evolving industry.

Case Study. Here, we discuss two examples of artists' careers: Jason Derulo and Anitta. They represent both markets, Global and Brazil, respectively. Table 5 summarizes the results for the selected artists for our discussion.

Jason Derulo is an American singer, songwriter, and dancer who rose to fame in the late 2000s. He starts his musical career as a songwriter, producing tracks for prominent artists like Diddy, Lil Wayne, and Sean Kingston. According to our data, Derulo explored more topics before and during his period of Hot Streak (i.e., his highest success already registered). The entropy for both periods is 1.0, representing *Exploration* phase. The respective songs for this period include *If I'm Lucky* and *Tip Toe*. The resulting genres by our genre network (Section 4) for these songs are *dance pop, eletro* and *rap, pop*. We search the respective genres over the internet (Wikipedia) to better check if the resulting genres conform to the real ones. The returned genres for *If I'm Lucky* and *Tip Toe* is *pop*, agreeing with our results. After the Hot Streak period of such an artist, his associated entropy decreases to 0.36, which indicates an *Exploitation* phase. In this case, the bet is to focus on fewer topics. In fact, songs for the period, such as *Goodbye* and *1, 2, 3*, are more like *latin* rhythms and *reggaeton*. Besides, one of his latest releases is the single *Slow Low* that is classified as *pop* and *reggaeton*. Still, our results are based on a particular time frame and may not fully capture the entirety of his career. External factors such as marketing strategies and industry trends could also influence the observed patterns. Further research and a broader dataset would be valuable for a more comprehensive understanding of Jason Derulo's artistic evolution.

Anitta is a Brazilian singer, songwriter and entrepreneur. She gained widespread recognition in Brazil and internationally for her unique blend of pop, funk, and reggaeton music styles. According to our results in Table 5, Anitta is fluctuating between *Exploration* and *Exploitation* in her phases (before, during, and after hot streaks). Specifically, during her hot streak, she had the most value for entropy result (0.59) – *Exploration* tendency. Despite that, she has most of her work transitioning between *pop* and *reggaeton* genres – by our methodology and even Wikipedia. Nevertheless, we believe she explored other topics (genres) before her hot streaks. After all, her career starts mainly as a *funk* singer (or *gospel*, at her church). However, our data comprise her career only from 2017 on, capturing initial results from her investments in the international recognition.

Table 5. Entropy results and their respective songs and genres by Hot Streak phases for Jason Derulo and Anitta.

Global - Jason Derulo				
Hot Streak Phase	**Entropy**	**Songs**	**Genres By Wikipedia**	**Genres By Community**
Before	1.0	If I'm Lucky / Swalla	[pop] / [r&b, dancehal]	[dance pop, eletro] / [r&b, rap]
During	1.0	Tip Toe	[pop]	[rap, pop]
After	0.36	Goodbye / 1,2,3	[latin, eletro] / [latin]	[dance pop, eletro] / [reggaeton, latin]
Brazilian - Anitta				
Hot Streak Phase	**Entropy**	**Songs**	**Genres By Wikipedia**	**Genres By Community**
Before	0.48	Paradinha / Sua Cara	[dancehall, reggaeton] / [Moombahton]	[dance pop, eletro] / [dance pop, eletro]
During	0.59	Machika, Indecente	[pop latino, reggaeton] / [pop latino]	[reggaeton, latin] / [dance pop, eletro]
After	0.45	Boys Don't Cry / Envolver	[electropop, pop rock] / [reggaeton]	[dance pop, eletro] / [dance pop, eletro]

6. Concluding Remarks

The music industry (like any other industry) deals with the constant natural evolution of the market, which reflects the need for adaptation to satisfy consumers. Trying to accommodate itself to potential changes, the market may explore the cycles of topics (e.g., musical genres) that are either up or down, where machine learning algorithms could be able to capture the nuances of when a cycle is about to end, and which one is predisposed to start by analyzing the available data. We know that understanding the behavior of creative careers (e.g., musical) is not a simple task due to their subjective nature (which considers factors that are not precisely measurable, such as creative capacity, personality, and other abilities). In addition, we must consider out-of-personal factors, such as the publicity force, social media power, etc. However, promoting a broader analysis of what can lead to a successful music career is still a necessary tool for the industry.

This paper analyzed how to identify regularities concerning the onset of hot streaks in music careers through a data-driven approach. Using a song network, we proposed a methodology for extracting relevant topics associated with such careers. Then, we used a method for measuring the entropy degree on such topics to quantify the level of exploration and exploitation about the first hot streak of every musical career. Our results are promising, showing there may be some regularity in prospecting topics to reach hot streaks. Specifically, artists explore diverse topics before hitting their first Hot Streak and during their most successful period. Then, we detected that artists tend to enter the exploitation phase as they leave their Hot Streaks periods. That is, they reduce the range of topics explored and are prone to specialize in fewer topics. Overall, our results highlight the important role of exploration and exploitation in individual careers, suggesting that a sequential view of such strategies balancing experimentation and implementation may be compelling for producing lasting careers. These findings could be relevant for identifying, training, and bringing up creative talent.

Limitations and Future Work. The data-driven nature of our study leads to common limitations in this type of analysis. First, regarding the datasets used to represent large collections of music career histories: they are limited to individuals who have had sufficiently long careers to provide enough data points for statistical analysis. Second, this paper presented correlational evidence, whose main objective is to investigate empirical regularities associated with the appearance of hot streak marks. We plan to explore other methods to extract topics from artists' careers and then the entropy methodology again.

Acknowledgments. Work supported by CAPES, CNPq, and FAPEMIG Brazil.

References

Amorim, A. et al. (2022). Modelagem de tópicos em textos curtos: uma avaliação experimental. In *Anais do XXXVII SBBD*, pages 254–266, Búzios, RJ. SBC.

Baldo, F., Grando, J., Weege, K., and Bonassa, G. (2022). Adaptive fast xgboost for binary classification. In *Anais do XXXVII SBBD*, pages 13–25, Búzios, RJ. SBC.

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022.

Byrd, D. and Crawford, T. (2002). Problems of music information retrieval in the real world. *Information Processing & Management*, 38(2):249–272.

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

Črepinšek, M., Liu, S.-H., and Mernik, M. (2013). Exploration and exploitation in evolutionary algorithms: A survey. *ACM Computing Surveys (CSUR)*, 45(3):1–33.

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

Liu, L. et al. (2018). Hot streaks in artistic, cultural, and scientific careers. *Nature*, 559(7714):396–399.

Liu, L. et al. (2021). Understanding the onset of hot streaks across artistic, cultural, and scientific careers. *Nature Communications*, 12(1):5392.

Martín-Gutiérrez, D. et al. (2020). A multimodal end-to-end deep learning architecture for music popularity prediction. *IEEE Access*, 8:39361–39374.

Oliveira, G. P. et al. (2020). Detecting collaboration profiles in success-based music genre networks. In *ISMIR*, pages 726–732.

Pachet, F. (2011). 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.

Passman, D. (2019). *All You Need to Know About the Music Business: 10th Edition*. Simon & Schuster, New York.

Seufitelli, D. B., Oliveira, G. P., Silva, M. O., Barbosa, G. R. G., Melo, B. C., Botelho, J. E., Melo-Gomes, L. d., and Moro, M. M. (2022). From compact discs to streaming: A comparison of eras within the Brazilian market. *Revista Vórtex*, 10(1):1–28.

Silva, M. O., Rocha, L. M., and Moro, M. M. (2019). Collaboration Profiles and Their Impact on Musical Success. In *ACM SAC*, pages 2070–2077, Limassol, Cyprus.

Yucesoy, B. and Barabási, A.-L. (2016). Untangling performance from success. *EPJ Data Science*, 5(1):17.