Understanding Musical Success Beyond Hit Songs: Characterization and Analyses of Musical Careers

Danilo Boechat Seufitelli
daniloboechat@dcc.ufmg.br
Universidade Federal de Minas Gerais
Belo Horizonte, Brazil

Mirella M. Moro
mirella@dcc.ufmg.br
Universidade Federal de Minas Gerais
Belo Horizonte, Brazil

ABSTRACT

Streaming brought the high data availability over the web associated with music consumption and listener preference. With such data, we can extract relevant knowledge, such as what can lead some songs to success and others not. In this scenario, Hit Song Science emerged, an area of study focused on revealing the dynamics of success in the music industry. Collecting hits can lead artists to experience periods of success far beyond the "ordinary" periods known as Hot Streaks. In this sense, understanding the factors of how the different profiles of artists stand out and reach their most successful periods can be crucial for the music industry, which deals with the constant natural evolution of the market and needs to reinvent itself to satisfy the desires of its consumers: connect successful music and artists. Thus, this thesis aims to identify the characteristics that lead artists to reach their most successful periods.

KEYWORDS

Hit Song Science, Music Information Retrieval, Musical Success

1 INTRODUCTION

The music industry is a dynamic environment in which artists' success goes beyond individual *hits* and includes the entire trajectory of their careers. For Computing Science, the sheer volume of music content available on the Web through streaming services poses new, interesting challenges. For instance, Spotify provides volumes of data then creating brand-new opportunities for extracting knowledge and benefiting the whole music ecosystem. Examples of computing applications include: creating music [3], exploring collaborations between artists [13], reconstructing audio missing parts [11], making enriched datasets available [4], and studying genre fluctuations [1].

Still, with so many musicians, composers and interpreters being released each year, a fundamental question for Artists and Repertories (A&R) and music producers is: how to differentiate and identify a potentially successful career. Pinpointing success is a complex puzzle per se, with previous works analyzing successful careers through different professions and perspectives, such as those based on genres, style of scientific publications, film characteristics, visual elements related to works of art, and even luck [5–8, 12].

In such a broad context, this thesis is an instance of Web available data-driven research, where we dive into music industry data to study mechanisms behind artists' success. Specifically, this thesis addresses characterization and analyses of musical careers by investigating periods of exceptional success, known as *hot streaks* [2, 7, 9], and the variables that precede them. The central problem is understanding how different factors contribute to these periods, facing challenges such as data variability and the lack of systematic studies on the progression of musical careers.

Traditional approaches to understanding success in music (such as feature engineering and machine learning - ML models) may not capture entirely the nature of the music industry. Hence, this thesis takes a data-driven approach to understanding the mechanisms behind an artist's success. By analyzing the evolution of musical genres and the topics of an artist's career, we aim to identify patterns that lead to successful periods. This approach sheds light on the creative process in music and has implications for identifying and nurturing creative talent in the industry. This objective reflects three specific Research Questions (RQs) that lead to achieving the overall goal of our work as follows: ***RQ1*** How does current research deal with Hit Song Science?; ***RQ2*** How does the evolution of music consumption in Brazil affect the occurrence of hot streaks?; and ***RQ3*** Do artists explore different topics in their careers before reaching their first Hot Streak?

2 METHODOLOGY

One innovative feature of this thesis is comparing music consumption patterns across different musical eras (physical – mostly selling records and radio airtime; and digital – mostly streaming consumption). Regarding advances over related work, this thesis leans on a multidimensional approach, which combines analysis of artist profiles, consumption patterns, and machine learning methods, overcoming the limitations of previous studies focused on isolated aspects. Finally, we discuss how artists' careers may achieve their first potential success by understanding the topic distribution of such careers beyond ML algorithms with a set of input data.

Our solution involves building a network of songs for each artist and applying aggregation techniques to simplify complex time series, as Figure 1 summarizes. We use Spotify's API to collect stream data and analyze the evolution of artists' success over time. Hot streak periods are identified through an individual threshold based on success percentiles, allowing a fair comparison between artists of different popularity levels. Advanced data analysis tools address issues such as data variability and complexity of collaborations.

3 CONTRIBUTIONS

The key contributions of this thesis include: *(1)* Development of a taxonomy and generic flow for Hit Songs Science; *(2)* Analysis of the evolution of music consumption in Brazil, identifying specific

In: VI Concurso de Teses e Dissertações (CTD 2024). Anais Estendidos do XXX Simpósio Brasileiro de Sistemas Multimídia e Web (CTD'2024). Juiz de Fora/MG, Brazil. Porto Alegre: Brazilian Computer Society, 2024.

ISSN 2596-1683

Figure 1: Five main stages to identify music genres (topics) based on hot streaks.

patterns for physical and digital eras; *(3)* Creation of artist clusters (Spike Hit Artists, Big Hit Artists, Top Hit Artists) that describe different levels of success; *(4)* Proposal of a methodology to measure the entropy of artists' careers, identifying exploration and exploitation phases; e *(5)* Publication of datasets for replication and extension of studies: MGD+ Dataset – public available in Zenodo.

Given this thesis's three research questions, the achieved findings for each one are summarized as follows.

RQ1. We proposed a generic workflow for Hit Song Prediction, and we also suggested novel taxonomies for (*i*) success measures, (*ii*) features, and (*iii*) learning methods used to consolidate the existing knowledge in HSS. We concluded there is not one "feature" for an ideal hit song prediction model, as its performance depends on subjective decisions made in the analysis process.

RQ2. We performed a clustered success analysis to answer the following questions: *are artists' most successful periods clustered in time? How to detect the artists' most successful periods (Hot Streaks)?* First, we witnessed the Hot Streaks and then performed a cluster analysis to understand if it is possible to distinguish the artists by their level of success. Next, we analyzed the most popular genres in the Brazilian market in different eras. Finally, we also verified if the artists underwent their first period of Hot Streak at similar times. Our main conclusion is that Brazilian listeners listen to Brazilian artists as their first option, independent of the era.

RQ3. We investigated whether artists explore different topics regularly before experiencing their first hot streak (exploration) and whether they specialize in specific segments from that point onward (exploitation). Hence, we propose extracting relevant topics in artists' careers by modeling their songs as a similarity network. Next, we applied a network's community detection and LDA algorithms to extract the career topics and then calculate the topic distribution entropy to measure the variation of topics encountered over such careers. Finally, we correlated the hot tracks' timing with the artists' creative trajectories to verify changes in the work characteristics. We found that artists explored more topics before and during their hot streak periods, and then they entered the exploitation phase when they left their hot streak periods. Such findings are valid for both global and Brazilian markets.

4 CONCLUDING REMARKS

Can computing explain the success of music careers by analyzing features beyond the results of machine learning algorithms? Yes, and we do so through a web data-driven approach. By analyzing the evolution of musical genres and the themes of an artist's career, we identified regularities and patterns that lead to periods of success. This approach sheds light on the creative process in music and helps identify and nurture creative talent in the industry.

This thesis uncovered that Hit Song Science has been gaining ground as a legitimate scientific field since the seminal paper that challenged its validity as such [10]. In this sense, we analyzed the evolution of music consumption by comparing data from the physical and digital eras. In the physical age, artists had to fight to get ratings on Radio shows and TV. With these appearances, artists could sell more vinyl records, CDs and DVDs. Generalizing physical media as records, the most successful artists sold the most. In the digital age, sales are mainly accounted for by the number of streams. The more streams an artist has, the higher their success. Finally, we studied if there is a certain regularity in terms of Exploration and Exploitation of topics that characterize the artist's career. Hence, we proposed a data-driven approach to the artists' musical genres to associate them with their songs with topics (genres) associated with each song, we can then build a timeline and detect the level of exploration and exploitation of such topics in artistic careers.

ACKNOWLEDGMENTS

This work was supported by CNPq, CAPES and FAPEMIG, Brazil.

REFERENCES

[1] Carlos V. S. Araujo and Eduardo F. Nakamura. 2018. Identification of Most Popular Musical Genres and their Influence Factors. In *Procs Braz Symp Multimedia and the Web*. ACM, Salvador, Brazil, 233–236. https://doi.org/10.1145/3243082.3264665
[2] Gabriel Barbosa, Bruna Melo, Gabriel Oliveira, Mariana Silva, Danilo Seufitelli, and Mirella Moro. 2021. Hot Streaks in the Brazilian Music Market: A Comparison Between Physical and Digital Eras. In *Anais do XVIII Simpósio Brasileiro de Computação Musical*. SBC, 152–159. https://doi.org/10.5753/sbcm.2021.19440
[3] Alexandra Covaci. 2023. Bytes and Rhythms – Delving into the Tech-Driven Evolution of Food and Music. In *Procs Braz Symp Multimedia and the Web*. ACM, Ribeirão Preto, Brazil, 1–2. https://doi.org/10.1145/3617023.3617024
[4] Angelo Cesar Mendes da Silva, Diego Furtado Silva, and Ricardo Marcondes Marcacini. 2020. 4MuLA: A Multitask, Multimodal, and Multilingual Dataset of Music Lyrics and Audio Features. In *Procs Braz Symp Multimedia and the Web*. ACM, São Luís, Brazil, 145–148. https://doi.org/10.1145/3428658.3431089
[5] Darryll Hendricks, Jayendu Patel, and Richard Zeckhauser. 1993. Hot hands in mutual funds: Short-run persistence of relative performance, 1974–1988. *The Journal of finance* 48, 1 (1993), 93–130.
[6] Milan Janosov, Federico Battiston, and Roberta Sinatra. 2020. Success and luck in creative careers. *The European Physical Journal Data Science* 9, 1 (2020), 9. https://doi.org/10.1140/epjds/s13688-020-00227-w
[7] Lu Liu et al. 2018. Hot streaks in artistic, cultural, and scientific careers. *Nature* 559, 7714 (2018), 396–399. https://doi.org/10.1038/s41586-018-0315-8
[8] Lu Liu, Nima Dehmamy, Jillian Chown, C. Lee Giles, and Dashun Wang. 2021. Understanding the onset of hot streaks across artistic, cultural, and scientific careers. *Nature Communications* 12, 1 (Sept. 2021), 5392. https://doi.org/10.1038/s41467-021-25477-8
[9] Gabriel P. Oliveira, Mariana O. Silva, Danilo B. Seufitelli, Gabriel R. G. Barbosa, Bruna C. Melo, and Mirella M. Moro. 2023. Hot streaks in the music industry: identifying and characterizing above-average success periods in artists' careers. *Scientometrics* 128, 11 (2023), 6029–6046. https://doi.org/10.1007/s11192-023-04835-x
[10] François Pachet and Pierre Roy. 2008. Hit Song Science Is Not Yet a Science. In *International Conference on Music Information Retrieval*. ISMIR, Philadelphia, USA, 355–360.
[11] Arthur C. Serra, Antonio José G. Busson, Álan L. V. Guedes, and Sérgio Colcher. 2021. Quality Enhancement of Highly Degraded Music Using Deep Learning-Based Prediction Models for Lost Frequencies. In *Procs Braz Symp Multimedia and the Web*. ACM, Belo Horizonte, Brazil, 205–211. https://doi.org/10.1145/3470482.3479635
[12] Danilo B. Seufitelli, Gabriel P. Oliveira, Mariana O. Silva, Gabriel R. G. Barbosa, Bruna C. Melo, Juliana E. Botelho, Luiza de Melo-Gomes, and Mirella M. Moro. 2022. From Compact Discs to Streaming: A Comparison of Eras within the Brazilian Market. *Revista Vórtex* 10, 1 (2022).
[13] Mariana O. Silva, Gabriel P. Oliveira, Danilo B. Seufitelli, Anisio Lacerda, and Mirella M. Moro. 2022. Collaboration as a Driving Factor for Hit Song Classification. In *Procs Braz Symp Multimedia and the Web*. Curitiba, Brazil, 66–74. https://doi.org/10.1145/3539637.3556993