

# Collaboration-Aware Hit Song Analysis and Prediction

Mariana O. Silva  
mariana.santos@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

We propose tackling the *Hit Song Prediction* problem through a multimodal form with songs' features fused together. Specifically, we describe songs from three feature modalities: *music*, *artist* and *album*. Initially, we identify collaboration profiles in a success-based musical network, unveiling how professional connections can significantly impact their success. Then, we use time series and the Granger Causality test for assessing whether there is a causal relationship between collaboration profiles and artists' popularity. Finally, we model the *Hit Song Prediction* problem as two distinct tasks: *classification* and *placement*. The former is a classical binary classification model and directly applies our fusion strategies. The latter is a modeling approach that ranks a song relative to a given chart, predicts hit songs, and provides comparative popularity information of a set of songs. Furthermore, we emphasize collaboration artists' profiles as important features when describing their songs. Overall, our empirical studies confirm the effectiveness of our method that fuses heterogeneous data for both tasks.

## KEYWORDS

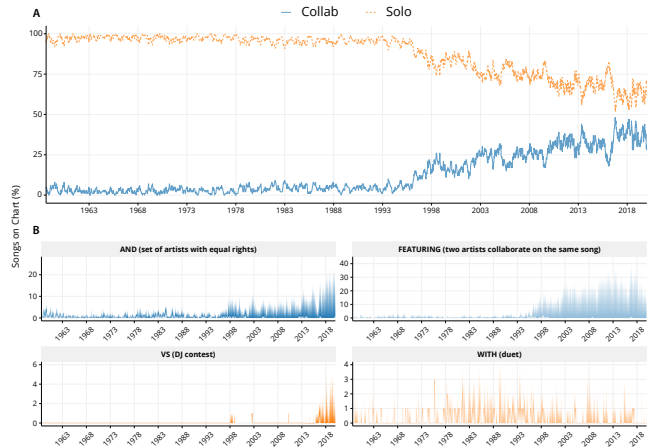
Collaboration Profiles, Hit Song Science, Musical Success, Social Networks, Machine Learning, Multimodal

## 1 INTRODUCTION

Usually, scientific research starts from a hypothesis and goes to the data to prove/disprove it. Data-driven research adds a new spin to research methodology: it starts from the data to come up with one (or more) hypothesis or to identify issues that need further investigation. This work is an instance of the latter with yet a new twist that we call **Web-available** Data-driven research, where we look into the music industry data available at the Web, on how hit songs behave in such industry and make up a hypothesis: collaboration is a relevant feature to hit songs and may even help predict them.

### 1.1 Problem Characterization

Hit songs establish the majority of a record company's profits; hence, labels invest billions into finding talent in the hope of hitting the jackpot. Creating a reliable scientific measure to predict whether a song offers the potential to become popular and commercially successful is a powerful and lucrative endeavor. One of the earliest studies on Hit Song Prediction [5] focuses on extracting acoustic and lyrical features from songs and using standard classifiers to separate hits from non-hits. The purpose was to determine if such



**Figure 1: Billboard Hot 100 songs (1958 - 2020). (A) The dashed line represents hits without collaborations, while the solid one reflects musical collaborations. (B) Pairwise historical frequency of collaboration types.**

a task was feasible or if the claims about the songs considered hits could be confirmed. A few years later, Pachet and Roy [15] stated that some subjective categories (such as style and character mood of the song) could be reasonably learned through techniques, but not popularity. However, after such a study, many authors persisted in discussing the feasibility of predicting a song's popularity by considering a suitable set of attributes. In other words, the secret to solving such a challenging task is on the data [4, 17].

Specifically, existing research in the area agrees that, besides the complex features to be measured, there are quantifiable qualities that contribute to the popularity of a song (see the dissertation Related Work section). Nonetheless, such current works ignore a feature that came up in our Web-available data analyses: how artists connect professionally. Indeed, artist collaboration is a strong force driving music today, as digital media enables musical collaboration among various artists. According to data from the Billboard Hot 100, the number of songs executed by more than one artist has significantly increased in recent years, as depicted in Figure 1A. In 2019, this number reached 42% of songs on the chart. There is a recent trend in increasing even more in the upcoming years.

Our data analysis was precise by showing that during the 1960s, 70s, and 80s, collaborations were almost rare (about 5% of charted songs) and generally took the form of duets (Figure 1B). The boom started in the mid-1990s, when the number of collaborations increased significantly, with the duets dying and the artists on display taking over. Such associations help artists bridge the gap between styles and genres, then crossing over to new fan bases. Nonetheless, factors driving the success of a collaborative process are not entirely understood. Also, for the sake of increasing market opportunities,

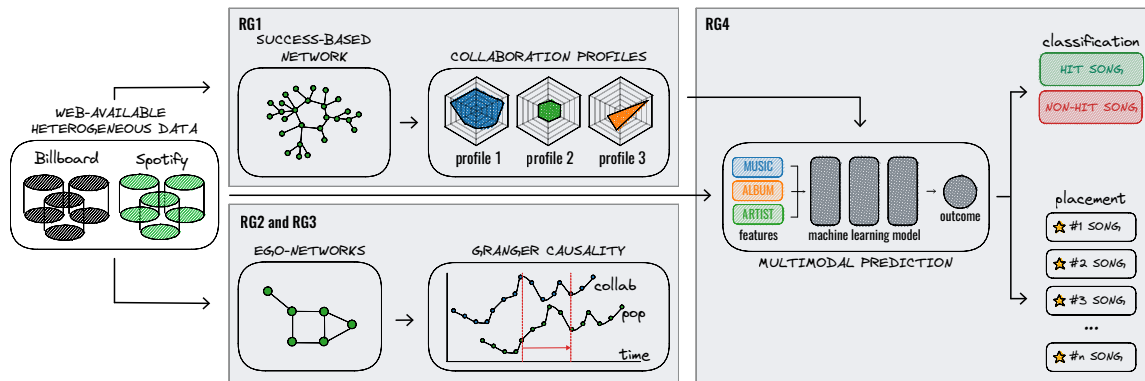


Figure 2: Overview of this work, according to the Research Goals (RGs).

investigating how music collaboration profiles can positively impact an artist’s popularity is still worth pursuing. For instance, while talent and status attract social connections, the researchers usually ignore that social networks can independently promote success [9]. One probable reason is the massive volume of Web data involved and the challenges such volume impose on any research.

## 1.2 Motivation

Music involves many features related to composition (melody, harmony, rhythm, lyrics) and social context (reachability and style of the artist, collaboration profiles, culture, and so on). With so many variables at hand, predicting song popularity remains a complex task requiring a balance between different musical feature sources. Learning from heterogeneous information may enable identifying correspondence between different features and gain an in-depth understanding of musical success.

Indeed, the ability to predict musical success offers huge benefits for many domains and audiences. For music industry CEOs, it may help maximize expected success by helping to decide whom to invest in to produce potential hits. Also, by properly investing in potential artist/music and distribution, the studio could increase physical and digital albums sales, improve revenue from on-demand audio streaming services, or even launch the next *pop star* or *Summer hit*. Artists may also profit by identifying the most suitable songs to lead the album to early stardom. For music consumers, it may help to decide if an album is worth buying because it may contain three to five hits instead of *one-hit* only.

## 1.3 Goal and Contributions

The goal of this work comes from Web-available data-driven research that revealed the high presence of collaboration in hit songs. From such a finding, we have defined the objective: evaluating the relevance of collaboration as a feature in hit songs and its potential power in predicting them. Specifically, we get data from the Web, organize it in social networks and apply machine learning techniques to exploit collaboration artists’ profiles as important features when describing and predicting hit songs. The general objective is then divided into four goals, achieved as follows (Figure 2).

**Research Goal 1 (RG1).** Identify the (potentially) topological measures and indicators that influence the popularity of both songs and artists (Dissertation Chapter 3). *Methodology summary:* build

a dataset from Web-available data sources (Billboard’s Artist 100 and Spotify); define successful artists; model social networks from artists collaborations; perform semantic characterization to group similar topological metrics into categories; and profile results using cluster and statistical analyses.

**Research Goal 2 (RG2).** Investigate the impact of these features on popularity over time, i.e., dynamically analyze whether features affect the popularity of an artist/song (Dissertation Chapter 4). *Methodology summary:* build dataset based on Web data; use ego-networks to model artists’ collaborations; define temporal success measures; and create three time series of musical success, musical collaborations and solo songs of the selected artists.

**Research Goal 3 (RG3).** Verify the causal relationship between collaboration profiles and music success (Dissertation Chapter 4). *Methodology summary:* build upon RG2 to perform Granger causality analysis.

**Research Goal 4 (RG4).** Propose a machine learning approach to derive a song’s popularity based on feature groups and identify the best way for combining them to predict song success (Dissertation Chapter 5). *Methodology summary:* to the best of our knowledge, we are the first to define Hit Song Prediction as two independent tasks – *binary classification* and *hit song placement*; then, we model it as a multimodal machine learning problem by considering three modalities – *music*, *artist* and *album*; we evaluate the performance of our proposed representations and interpret the learned models to identify the most important features.

## 2 RESULTS AND DISCUSSION

This dissertation explores the relationship between collaboration profiles and musical success towards predicting whether a given song will rise to a high position in top charts. We define such analyses into research goals, discussed next. In addition, we cite the publications that are directly and/or indirectly related to each research goal (all of them were published during the Master’s program or submitted shortly afterward).

**RG1: Identify the (potentially) intrinsic features and indicators that influence the popularity of both songs and artists** [27]. We analyzed collaborations in a musical success-based network built from Web data. Through six topological metrics, we

**Table 1: A brief comparative analysis of existing Hit Song Science research studies**

Task	Musical Features	Social Information	Year	Reference
Regression	tweets	listeners' information	2017	[1]
Regression	acoustic features	<i>none</i>	2017	[28]
Classification	acoustic features and artists' reputation	artists' reputation	2018	[6]
Classification	acoustic features	<i>none</i>	2018	[7]
Classification & Regression	user songs operations and relations between songs and artists	listeners' information	2019	[29]
Classification	popularity information and acoustic features	<i>none</i>	2019	[2]
Regression	low- or high-level audio features	<i>none</i>	2019	[30]
Classification & Regression	text, audio and metadata	<i>none</i>	2020	[8]
<b>Classification &amp; Placement</b>	<b>music-, artist- and album- related features</b>	<b>collaboration profiles and artists' metadata</b>	<b>2020</b>	<b>Self</b>

defined four collaboration profiles: *Interaction*, *Distance*, *Influence*, and *Similarity*. The first three affect musical success more intensely than *Similarity* i.e., successful artists are more likely to have a high degree of collaboration between influential and diversified artists. Such findings suggest there is a powerful sense that collaborating enhances artists' prospects of having a successful song, not only boosting unknown artists but also reinvigorating careers of more established ones. Nevertheless, those who prefer to pursue a non-collaborative musical career may be missing an opportunity to enhance and expand their potential. Our results provide strong evidence that (i) there are, in fact, distinct success factors for musical collaboration profiles that are socially measurable, and (ii) there are common factors for successful collaboration in the music market. Overall, our analysis motivates the further study of the causal relationship between such profiles and musical success.

**RG2 and RG3: Investigate the impact of these features on popularity over time, and verify the causal relationship between collaboration profiles and music success [24].** From RG1, we proposed to establish relations of causality between collaboration profiles and artists' popularity. Previous findings reveal that the way artists connect professionally may affect their musical success. Hence, we further such analyses by using time series and the Granger causality test for assessing whether there is a causal relationship between collaboration profiles and artistic success. Our evaluation reveals varied relationships linking collaboration profiles and musical success, indicating a direct connection. In particular, we identified the presence of a cycle between collaborations and musical success: before collaborating with others, most artists may consider the level of success of their collaborators. It means collaboration can be seen as a musical career propeller, working as a fast-track route to becoming a household name, especially when reaching out to more established artists. Furthermore, based on findings regarding RG1, we found that the few collaborations performed by non-collaborative artists exposed a substantial impact on their success. However, such artists are part of the group with the lowest average success rate. This indicates that, because they collaborate very little, they may be missing the opportunity to grow their reach and inspire themselves to new creative heights. Whereas our results showed an unclear causal relationship between collaboration profiles and musical success, these findings generally offer a novel perspective on success in the music industry, unraveling how collaboration profiles can contribute to an artist's popularity.

**RG4: Propose a machine learning approach to derive a song's popularity based on these groups of features and determine**

**the best way for combining them to predict the success of a song [20, 26].** As a result of the collaboration profiles-musical success correlation, a natural next step was to propose a methodology for predicting hit songs. Hence, we tackle Hit Song Prediction from two distinct tasks: *binary classification* and *placement*. To assess both tasks, we model a song as a multimodal representation using information from three modalities: *music*, *artist* and *album*. To verify our results, we focused mainly on artistic collaboration features extracted from Web-available data. By comparing the performance of models with different fusion strategies (considering different musical modalities), we found that relying exclusively on internal musical characteristics is not enough to obtain correct hit song predictions—moreover, a feature importance analysis allowed to identify the most significant features that drive hits prediction. We found the *artist* modality contains the most significant predictors, mainly social interaction information. Such results demonstrate the relevance of handling Hit Song Prediction as a multimodal problem and the importance of relying on information from the artists' collaboration profiles. To sum up, our results reveal it is possible to successfully predict whether a given song will be a hit or not.

**Novelty Insights.** In general, approaches that consider only acoustic features and/or lyric-based information do not imply a proper musical representation for the prediction task (see the dissertation Section 2.1.1 Acoustic and Lyric-Based Features). Such unimodal representations disregard how external factors influence a song's popularity, including the artists' reputation, social influence, historical context, or any other extrinsic feature that led a song to have a peak of success and become a hit. Further studies investigated music success with a joint representation by including non-musical factors and the relationship between musical and social information, e.g. [1] (see the dissertation Section 2.1.2 Listener's Information). Mostly, they found that the prediction accuracy improves when incorporating external factors. However, they generally focus on specific social attributes which are limited (Table 1).

Our work differs fundamentally from existing works, as we introduce collaboration-aware modalities towards hit song prediction. Moreover, most of the proposed works on HSS are limited to considering specific contexts (a given genre or a certain period), focusing on one or two attributes (acoustics and/or on lyric content), or relying on data of a particular social network. Hence, there is a strong potential for modeling music success through a multimodal and suitable combination of heterogeneous data. In this sense, our study is innovative because it proposes merging the music content and subjective social attributes and collaboration influence to predict

success. As shown in Table 1, we are the first authors to consider collaboration profiles as predictors to assess the Hit Song Prediction problem. Furthermore, our work is the first to employ the Learning to Place approach in the HSS context. Finally, this whole research was inspired by and relied solely on data available on the Web, which we organized in proper datasets [14, 21, 25].

### 3 RESEARCH ACCOMPLISHMENTS

**Relevance Insights.** Our work differs from the state-of-the-art in two crucial ways. First, although considering social aspects in hit song prediction is beneficial from the analytical perspective, this is the first time that the *collaboration between artists and their profiles* are modeled as features for a machine learning approach. Second, the multimodal perspective brings the necessary complexity to analyzing music in all its facets. Therefore, combining the multimodal representation with a collaboration-aware model means a big step towards advancing both fields of *Hit Song Science* and *Multimodal Music Analysis*, providing potential impact to the music industry and multimedia systems that consider music files as well.

**Products.** The research developed in this dissertation has contributed to the following products:

- study on music collaboration profiles in a musical success-based network built from Web-available data [27];
- analyses on the relations of causality between collaboration profiles and artists' popularity [24];
- multimodal methodology to predict hit songs [26];
- MUSICOSET [20, 21], enhanced dataset of musical elements (artists, songs and albums) based on musical popularity;
- COMUSIC [25], dataset with a successful musical collaboration network, billboard charts, 30 ego networks, and time series; and
- MGD [14], music genre dataset with genre networks and mapping, artist networks and mapping, hit songs and charts.

**Byproducts.** This dissertation has also inspired further research in which we have worked with our group:

- teaching material that uses our datasets, published in a book chapter and presented during the CSBC 2021 [16];
- an improvement to detect collaboration profiles by considering genre [11] (this publication has won Best Poster Presentation Sessions 4-6 at ISMIR 2020);
- an extensive survey on the hit song science [18];
- opportunities to extend the contributions of the dissertation to specific problems such as hot streak [3, 10, 12, 13]; and
- knowledge to apply concepts from the dissertation into another context, Literature [19, 22, 23] (these publications have won Best Paper Runner Up at BraSNAM 2021 and Best Paper at DSW 2021).

### ACKNOWLEDGMENTS

The first author thanks CNPq for her Masters Scholarship.

### REFERENCES

- [1] Carlos V. S. Araújo et al. 2017. Predicting Music Success Based on Users' Comments on Online Social Networks. In *Braz Symp on Multimedia and the Web - WebMedia*. 149–156. <https://doi.org/10.1145/3126858.3126885>
- [2] Carlos Vicente Soares Araujo et al. 2019. Predicting Music Popularity Using Music Charts. In *ICMLA*. 859–864. <https://doi.org/10.1109/ICMLA.2019.00149>
- [3] Gabriel R. G. Barbosa, Bruna C. Melo, Gabriel P. Oliveira, Mariana O. Silva, Danilo B. Seufitelli, and Mirella M. Moro. 2021. Hot Streaks in the Brazilian Music Market: A Comparison Between Physical and Digital Eras. In *Simp. Bras. Computação e Música*.
- [4] Kerstin Bischoff et al. 2009. Social Knowledge-Driven Music Hit Prediction. In *ADMA*. 43–54. [https://doi.org/10.1007/978-3-642-03348-3\\_8](https://doi.org/10.1007/978-3-642-03348-3_8)
- [5] Ruth Dhanaraj and Beth Logan. 2005. Automatic Prediction of Hit Songs. In *Proc Int'l Conf on Music Information Retrieval, ISMIR*. 488–491.
- [6] Myra Interiano et al. 2018. Musical trends and predictability of success in contemporary songs in and out of the top charts. *Royal Society open science* 5, 5 (2018), 171274.
- [7] Junghyuk Lee and Jong-Seok Lee. 2018. Music Popularity: Metrics, Characteristics, and Audio-Based Prediction. *IEEE Trans. Multimedia* 20, 11 (2018), 3173–3182. <https://doi.org/10.1109/TMM.2018.2820903>
- [8] D. Martín-Gutiérrez et al. 2020. A Multimodal End-to-End Deep Learning Architecture for Music Popularity Prediction. *IEEE Access* 8 (2020), 39361–39374. <https://doi.org/10.1109/ACCESS.2020.2976033>
- [9] Siobhan McAndrew and Martin Everett. 2015. Music as Collective Invention: A Social Network Analysis of Composers. *Cultural Sociology* 9, 1 (2015), 56–80. <https://doi.org/10.1177/1749975514542486>
- [10] Gabriel Oliveira, Gabriel R. G. Barbosa, Bruna C. Melo, Mariana O. Silva, Danilo Seufitelli, and Mirella M. Moro. 2021. MUHSIC: An Open Dataset with Temporal Musical Success Information. In *SBB: Dataset Showcase Workshop*.
- [11] Gabriel Oliveira, Mariana Santos, Danilo Seufitelli, Anísio Lacerda, and Mirella M. Moro. 2020. Detecting Collaboration Profiles in Success-based Music Genre Networks. In *ISMIR*. 726–732.
- [12] Gabriel P. Oliveira, Gabriel R. G. Barbosa, Bruna C. Melo, Mariana O. Silva, Danilo B. Seufitelli, Anísio Lacerda, and Mirella M. Moro. 2021. Hot Streaks in Musical Careers. *WWWJ* (2021). [under submission].
- [13] Gabriel P. Oliveira, Gabriel R. G. Barbosa, Bruna C. Melo, Mariana O. Silva, Danilo B. Seufitelli, Anísio Lacerda, and Mirella M. Moro. 2021. MUHSIC: An Open Dataset with Temporal Musical Success Information. *Zenodo*. <https://doi.org/10.5281/zenodo.5168695>
- [14] Gabriel P. Oliveira, Mariana O. Silva, Danilo B. Seufitelli, Anísio Lacerda, and Mirella M. Moro. 2020. MGD: Music Genre Dataset. *Zenodo*. <https://doi.org/10.5281/zenodo.4778563>
- [15] François Pachet and Pierre Roy. 2008. Hit Song Science Is Not Yet a Science. In *Proc Int'l Conf on Music Information Retrieval, ISMIR*. 355–360.
- [16] João Felipe Pimentel, Gabriel P. Oliveira, Mariana O. Silva, Danilo B. Seufitelli, and Mirella M. Moro. 2021. Ciência de Dados com Reprodutibilidade usando Jupyter. In *Jornada de Atualização em Informática 2021*, Aline Maria Santos Andrade and Raul Sidnei Wazlawick (Eds.). SBC, Chapter 1, 13–62. <https://doi.org/10.5753/sbc.6757.3.1>
- [17] Jing Ren, Jialie Shen, and Robert J. Kauffman. 2016. What Makes a Music Track Popular in Online Social Networks?. In *Proc Int'l Conf World Wide Web - Companion Volume*. <https://doi.org/10.1145/2872518.2889402>
- [18] Clarisse Scofield, Gabriel P. Oliveira, Mariana O. Silva, Danilo B. Seufitelli, and Mirella M. Moro. 2021. Hit Song Science: A Comprehensive Survey and Research Directions. (2021). [under submission].
- [19] Mariana Silva et al. 2021. Exploring Brazilian Cultural Identity Through Reading Preferences. In *BraSNAM*. 115–126. <https://doi.org/10.5753/brasnam.2021.16130>
- [20] M.O. Silva, Laís Rocha, and M.M. Moro. 2019. MusicOSet: An Enhanced Open Dataset for Music Data Mining. In *SBB: Dataset Showcase Workshop*.
- [21] Mariana O. Silva et al. 2019. MusicOSet: An Enhanced Open Dataset for Music Data Mining. *Zenodo*. <https://doi.org/10.5281/zenodo.4904639>
- [22] Mariana O. Silva et al. 2021. PPORTAL: Public Domain Portuguese-language Literature Dataset. In *SBB: Dataset Showcase Workshop*.
- [23] Mariana O. Silva and Clarisse Scofield Mirella M. Moro. 2021. PPORTAL: Public domain Portuguese-language literature Dataset. *Zenodo*. <https://doi.org/10.5281/zenodo.5178063>
- [24] Mariana O. Silva and Mirella M. Moro. 2019. Causality Analysis Between Collaboration Profiles and Musical Success. In *Braz Symp on Multimedia and the Web - WebMedia*. 369–376. <https://doi.org/10.1145/3323503.3349549>
- [25] Mariana O. Silva, Laís Mota, and Mirella M. Moro. 2019. CoMusic: Good things come to those who collaborate. *Zenodo*. <https://doi.org/10.5281/zenodo.4904676>
- [26] Mariana O. Silva, Gabriel P. Oliveira, Anísio Lacerda, and Mirella M. Moro. 2021. Collaboration-Aware Multimodal Hit Song Prediction. *Information Processing and Management* (2021). [under submission].
- [27] Mariana O. Silva, Laís M. Rocha, and Mirella M. Moro. 2019. Collaboration Profiles and Their Impact on Musical Success. In *ACM/SIGAPP SAC*. <https://doi.org/10.1145/3297280.3297483>
- [28] Li-Chia Yang et al. 2017. Revisiting the problem of audio-based hit song prediction using convolutional neural networks. In *ICASSP*. <https://doi.org/10.1109/ICASSP.2017.7952230>
- [29] Haiqing Yu et al. 2019. Popularity Prediction for Artists Based on User Songs Dataset. In *ICCAI* 17–24. <https://doi.org/10.1145/3330482.3330493>
- [30] Eva Zangerle et al. 2019. Hit Song Prediction: Leveraging Low- and High-Level Audio Features. In *ISMIR*. 319–326.