

ORIGINAL ARTICLE

A quantitative comparison of viral and hit songs in the Brazilian music market

Gabriel P. Oliveira 

Ana Paula Couto da Silva 

Mirella M. Moro 

Universidade Federal de Minas Gerais | Belo Horizonte, MG, Brazil

Resumo: A viralização de músicas através de plataformas de streaming e redes sociais é comum, mas nem todas as músicas virais se tornam sucessos. Neste contexto, nosso objetivo é descobrir o que difere as músicas virais dos hits para além da definição. Nós utilizamos uma metodologia quantitativa em paradas de sucesso do mercado brasileiro. Comparamos músicas de sucesso e virais quanto às suas características intrínsecas e extrínsecas, e os resultados revelam diferenças significativas entre elas. Características como gêneros musicais, tópicos das letras e emoções surgem como elementos cruciais para distinguir tais canções no contexto brasileiro. Além disso, características temporais indicam diferenças nos processos de difusão entre hits e virais. Em geral, este estudo oferece percepções sobre o consumo de música no Brasil, revelando a conexão entre as características das músicas e seu sucesso e viralização em plataformas de streaming.

Palavras-chave: viralização de músicas, sucesso musical, análise quantitativa, Brasil.

Abstract: It is common for songs to go viral on streaming platforms and social media, but not all viral songs become hits. In this context, we aim to discover what differs viral from hit songs beyond their definition. We do so by using a quantitative methodology over charts in the Brazilian market. We compare hit and viral songs regarding their intrinsic and extrinsic characteristics, and our results reveal significant differences between them. Features such as music genres, lyrics topics, and emotions emerge as crucial elements to distinguishing such songs within the Brazilian context. Furthermore, temporal features indicate differences in the diffusion processes between hits and virals. Overall, this study offers insights into music consumption in Brazil, revealing the connection between song features and their success and virality on streaming platforms.

Keywords: song virality, musical success, quantitative analysis, Brazil.

Every day, people have access to a massive volume of content on the Internet, especially on Social Networks. In such platforms, users can share and repost content (i.e., a blog post, a video, or a song) from others at any moment, and some posts get a lot of shares quickly, reaching several other users. In social media, “going viral” means that specific content spreads quickly across platforms, being shared by thousands or even millions of users in a short time span (Guerini; Strapparava; Özbal, 2011). Indeed, such processes are inherently social, meaning they heavily depend on people’s actions -- whether sharing content online or talking about it (Le Compte; Klug, 2021).

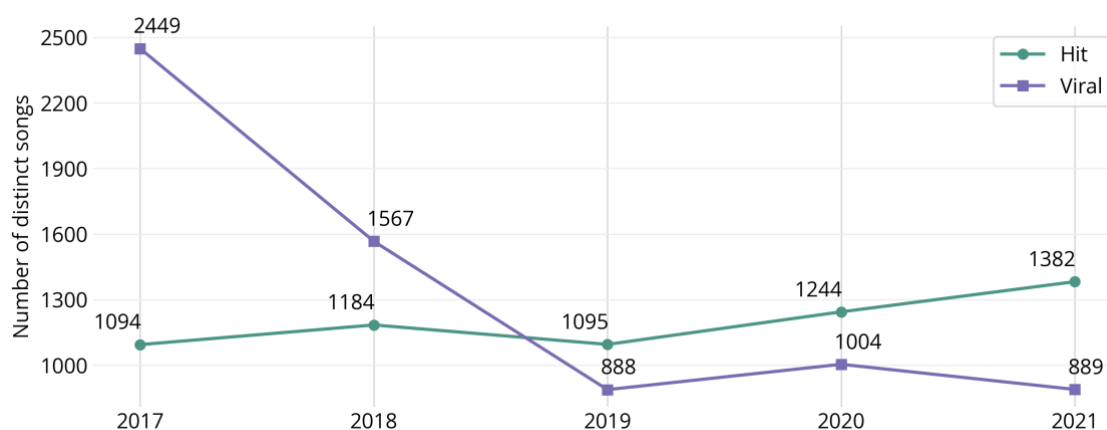
Streaming services are now the most used form of consuming music, and platforms such as Spotify and YouTube allow users to share what they are listening to with their contacts directly. According to the International Federation of the Phonographic Industry (IFPI), audio and video streaming collectively account for 62% of the time individuals dedicate to interacting with music in 2023 (International Federation of the Phonographic Industry, 2023). Such a relevance is also reflected in the economy. In Brazil, streaming is responsible for 86.2% of the total revenue of the national phonographic market in 2022 (Pró-Música Brasil, 2023). Indeed, the streaming ecosystem enables the viral spread of songs, but this process is also affected by other social platforms. For example, TikTok, a video-sharing platform, often contributes to songs going viral and becoming widely popular through the videos its users create.

Analyzing music popularity is not a trivial task since there is not a singular, universally accepted definition. It may be related to a high number of streams but also to an extensive discussion about a song (online or offline) (Seufitelli et al., 2023b). Regardless of the definition being used, a viral song does not necessarily mean that it is a hit. Both concepts are related to popularity but differ in how they are measured. Whereas virality relies on content sharing and is an ephemeral process, success is more solid and can be measured in several ways, including streams, sales, or radio airplay (Guerini; Strapparava; Özbal, 2011; Oliveira; Couto da Silva; Moro, 2024).

Besides, virality can be a stepping stone for musical success. A good example of a song that achieved high success after going viral is “Envolver” by Anitta. The song was originally released in 2021, but only in early 2022 it started to go viral because of its choreography. The song’s TikTok dance challenge (also known as “*El paso de Anitta*”) involves fans imitating a well-known dance move

of the singer, where she drops to the floor and moves rhythmically to the beat of the song. The virality reflected in success, as the song reached the top spot on the Spotify Global Chart on March 24, with over 6.3 million daily streams worldwide.¹

FIGURE 1 – Number of distinct songs in Spotify Charts in Brazil per year (2017-2021).



With new social platforms changing how people consume music, virality has taken a key role in music popularity. Following such a transformation, streaming platforms started producing distinct rankings for viral and hit (i.e., successful) songs. Specifically, considering charts from Spotify reveals different trends: in Brazil, the number of distinct viral songs decreases over time, while the number of hits increases (Figure 1). In this work, we use the top-charts definition for viral and hit songs to answer the following questions: **“Are viral and hit songs two sides of the same coin? What differs them in Brazil?”** Hence, we perform a comparative analysis of hit and viral songs consumed in Brazil (including songs from Brazilian and foreign artists), as extracted from their respective charts on Spotify, to unveil similar and distinct patterns in the Brazilian market.

We achieve such a goal through a quantitative and data-driven methodology, in which we analyze an enhanced set of song features extracted from Spotify and Genius, two of the most relevant online music-related platforms. First, we analyze the intrinsic features, which are directly derived from the songs. Then, we assess extrinsic features related to other aspects of songs, such as their artists

¹ Spotify Charts - March 24, 2022: <https://bit.ly/49an0nB>

and their behavior on the charts. The main contributions of this paper rely on structured and statistical analyses that reveal distinct characteristics of hit and viral songs, highlighting the complexity of the music consumption process.

The remainder of this article is organized as follows. First, we discuss related work in Section 1. Then, we describe our methodology in Section 2. Next, we analyze the intrinsic and extrinsic features in Sections 3 and 4, respectively. In Section 5, we present case studies to qualitatively illustrate our previous results, and we discuss all of them in Section 6. Finally, we present our concluding remarks in Section 7.

1. Related Work

The relationship between songs' features and their success has been extensively studied in Music Information Retrieval (MIR). From the seminal work of Dhanaraj and Logan (2005), several types of features (and their combinations) were considered to reach such a goal. Indeed, they were the pioneers in using acoustic and lyric-based features when looking for evidence that there is indeed a pattern connecting hit songs. Such features are still widely used in recent works (Araujo; Cristo; Giusti, 2019; Silva et al., 2022), as they act as descriptors of the core elements of a song. Furthermore, the dynamic nature of music and the emergence of social platforms have provided new features to assess song popularity, including social influence (Cosimato et al., 2019; Tsiara; Tjortjjs, 2020) and music collaboration (Oliveira et al., 2020; Silva et al., 2022).

Online platforms allow people to share content (including music) anytime, anywhere. In fact, their sharing nature enhances the viral phenomenon of their content. For example, YouTube was one of the first platforms to manifest such a phenomenon, and research on such a platform focuses mainly on predicting videos' popularity (Jiang et al., 2014; Kong et al., 2018). More recently, TikTok emerged as the main platform in which it is possible to observe content virality. Recent studies assess personal motivations and behaviors (Le Compte; Klug, 2021), and the content features that may be behind virality (Ling; De Cristofaro; Stringhini, 2022).

Previous research on song popularity revealed that analyzing regional markets individually is also critical, as each one behaves in its own way and has distinct patterns of music consumption (Oliveira et al., 2020; Vaz de Melo; Machado; Carvalho, 2020). For instance, Brazil is a continental country with a vibrant music scene, in which local genres and artists play a key role in viral and hit songs. Such an ecosystem is unique, and findings related to other music markets may not apply to the Brazilian context.

Specifically, Seufitelli et al. (2022) analyze musical artists' careers in Brazil to detect hot streak periods, i.e., periods in which the success is above the normal. Their findings evidence that the most successful artists in the Digital Era (i.e., when streaming became the most used form to consume music) are mostly Brazilians, indicating a strong preference for local artists and genres. Later on, Silva et al. (2023) show that artists and genres can present different collaboration patterns when comparing the Global and Brazilian music markets.

Regarding the conceptual differences between virality and success, little is known about how these two concepts relate in the music context. Therefore, we aim to evaluate and compare both viral and hit songs in Brazil. By examining how such processes shape the popularity of songs in this specific cultural context, our study contributes to a better understanding of the music consumption dynamics in the Brazilian music market.

2. Methodology

In this section, we first describe how we built the dataset of hit and viral songs consumed in Brazil. Then, we present the set of features applied to investigate potential differences among songs concerning success and virality.²

² Note that our intention is not to reverse engineer any existing ranking mechanism, nor do we aim to predict which songs will become hits or go viral. In fact, our goal is to characterize such songs to better understand the differences between hits and virals.

2.1. Data Collection

Our dataset consists of songs' information from two platforms: Spotify³ provides song charts and songs' metadata, and Genius⁴ shares songs' lyrics.

Spotify. As one of the world's most popular streaming services, Spotify has more than 551 million users in 184 markets and is the most used music streaming app in Brazil.⁵ Daily and weekly charts for **hit** and **viral** songs by country are provided by the platform. Specifically, we consider as hit and viral all songs present in the Top 200 and Viral 50 charts, respectively. In other words, the ranking of the hit songs includes the most listened-to (i.e., streamed) songs. The ranking of viral songs, instead, captures the songs gaining the most buzz on the platform by considering the rise in plays, the number of sharing, and the number of people who have recently discovered the song.⁶

Our Spotify dataset contains the daily charts of the top 200 hit songs and the top 50 viral songs in Brazil, enriched with their metadata (artists, acoustic features, release dates, and genres). The data collection spans from January 2017 to March 2022, acquired from the Music Genre Dataset (MGD+) (Seufitelli et al., 2023a) for hits and through the Spotify Web API⁷ for virals.

Genius. With a community of more than two million contributors, Genius is a collaborative platform for sharing music knowledge, enabling individuals to provide facts and insights about songs and artists. Users' contributions are moderated by a team of editors, ensuring the quality of the provided information. We collected the lyrics by scrapping the Genius website, and the metadata of both viral and hit songs through the Genius API.⁸

³ Spotify (Oct. 2023): <https://investors.spotify.com/about/>

⁴ Genius: <https://genius.com/>

⁵ According to the Panorama Mobile Time/Opinion Box Research. Available on Terra (Oct. 2023): <https://bit.ly/45sXoQ0>

⁶ Spotify: <https://bit.ly/3QfZ35W>

⁷ Spotify API: <https://developer.spotify.com/>

⁸ Genius API: <https://docs.genius.com/>

Overall, from Spotify's charts, we collected 5,010 distinct hit songs and 6,699 distinct viral songs (i.e., our full dataset). However, not all of these songs have their lyrics available on Genius. Therefore, we have a second lyrics-based dataset with 3,783 (75.5%) hit songs and 4,963 (74.1%) viral songs, respectively. We use the complete dataset in all our analyses, except for the lyrics-related features (Sections 2.2 and 3.3), in which we consider the reduced dataset.

2.2. Set of Features

To investigate whether there are specific features that distinguish viral from hit songs, we focus on two groups of features based on their relationship with the song, following the taxonomy proposed in Seufitelli et al. (2023b). The first group, called *intrinsic features*, considers the song's metadata, acoustic and lyrics-related features. The second group, called *extrinsic features*, focuses on artists and chart-related data. Next, we briefly describe such features.

TABLE 1 – Acoustic features considered in this work, with their types and descriptions as provided by Spotify.

Feature	Type	Description
acousticness	Float	Probability of a song being acoustic or not
danceability	Float	Probability of a song being suitable for dancing
duration_ms	Integer	Duration of a song in milliseconds
energy	Float	Intensity and activity of a song in terms of perceived loudness, timbre, and general entropy
instrumentalness	Float	Probability of a song being instrumental, i.e., without vocals
key	Integer	Estimated overall key of a song, mapped as an integer number (e.g., C=0, C#=1, and so on)
liveness	Float	Probability of a song being performed live, i.e., the presence of an audience in a song
loudness	Float	General loudness measured in decibels (dB)
speechiness	Float	Probability of a given song having spoken words in it
tempo	Float	Speed of the song, measured in beats per minute (BPM)
time_signature	Integer	Amount of beats in each bar (measure)
valence	Float	Positiveness of a song (in which high valence values represent happier songs, whereas low values, the opposite)

Metadata. Here, we select two types of information. The first one regards the relationships of song with respect of previously released songs, i.e., whether a song contains a *sample*, is a *remix* or a *cover*.⁹ Such versions are becoming widely used in viral videos on TikTok and Instagram.¹⁰

The second information we are interested in is the music genre. Since Spotify's genres are associated with artists rather than individual songs, we associate each song with all genres attributed to the artists who sing it. We acknowledge that musical genres are dynamic and constantly changing, with new genres constantly emerging from the combination or adaptation of existing genres. However, in this work we use the musical genres as provided by the Spotify API without performing any transformation or categorization.

Acoustic Features. Provided by the Spotify API, these features rely on musical data extracted from the audio properties (e.g., pitch, rhythm, dynamics, and timbre). Such features have been extensively studied and shown to associate with the motivations behind music listening, particularly regarding emotional coping mechanisms (Duman et al., 2022). Table 1 lists all features used in our analyses, with their respective types (i.e., whether they are represented by integer or decimal/floating-point numbers) and descriptions.

Lyrics-related Features. Lyrics are frequently used in music-related research for evaluating rhyme and text. We analyze four types of lyrics-related features: (i) *General characterization*, with the number of words, lines, and verses; (ii) *Language*, a categorical feature for the language of a song; (iii) *Main topics*, extracted by using the Latent Dirichlet Allocation (LDA) algorithm (Blei; Ng; Jordan, 2003); and (iv) *Psycholinguistic features*, extracted by using Linguistic Inquiry and Word Count (LIWC) (Tausczik; Pennebaker, 2010), which assigns words within a given text to linguistic and psychological dimensions (e.g., emotions, word categories, slangs, and so on).

⁹ In short, sampling involves taking a portion of an existing song and using it as part of a new one; remixing is altering a song to create a new version of it; and a cover is a new performance of an existing song. Such aspects are not mutually exclusive, as a song may present one or more of them at the same time (i.e., being a remix and containing a sample from another song).

¹⁰ NBC News: <https://nbcnews.to/47Q2YO1>

Artist-related Features. Analyzing the artists who sing a given song may reveal important dimensions of a song's virality and popularity. We focus on artist collaboration (i.e., when two or more artists are involved in a song), a specific dimension which previous studies directly relate to musical success (Bischoff et al., 2009; Silva et al., 2022).

Temporal Features. Regarding song popularity, research studies usually consider information such as position in charts to grasp the level of success. We consider Spotify's charts to derive two temporal features: the time from the songs' release until they reach the charts for the first time, and the time they spend on the charts. Both features are measured in days. Note that the period of the last feature is not necessarily continuous. For example, if a song stays ten days on the charts, leaves, and then re-enters for five days, the time it spends on the charts is 15 days.

TABLE 2 – Proportion of song relationships with other pre-existing songs.

	Hit		Viral	
	Songs	%	Songs	%
Sampling	342	9.04%	460	9.27%
Remix	190	5.02%	280	5.64%
Cover	81	2.14%	155	3.12%

TABLE 3 – Top 10 most frequent music genres of artists whose songs are hit or viral in Spotify Brazil.

Genre	Hit		Genre	Viral	
	Songs	%		Songs	%
pop	1,451	38.36%	pop	1,116	22.49%
sertanejo universitário	739	19.53%	rap	486	9.79%
arrocha	670	17.71%	funk carioca	450	9.07%
funk carioca	579	15.31%	pop nacional	374	7.54%
pop nacional	542	14.33%	hip hop	340	6.85%
sertanejo	503	13.30%	brazilian hip hop	335	6.75%
agronejo	337	8.91%	sertanejo universitário	302	6.09%
dance pop	325	8.59%	dance pop	292	5.88%
rap	319	8.43%	k-pop	266	5.36%
funk rj	255	6.74%	arrocha	251	5.06%

3. Intrinsic Features

This section delves into the analysis of the intrinsic features of the songs: metadata (Section 3.1), acoustic (Section 3.2), and lyrics-related (Section 3.3).

3.1. Metadata

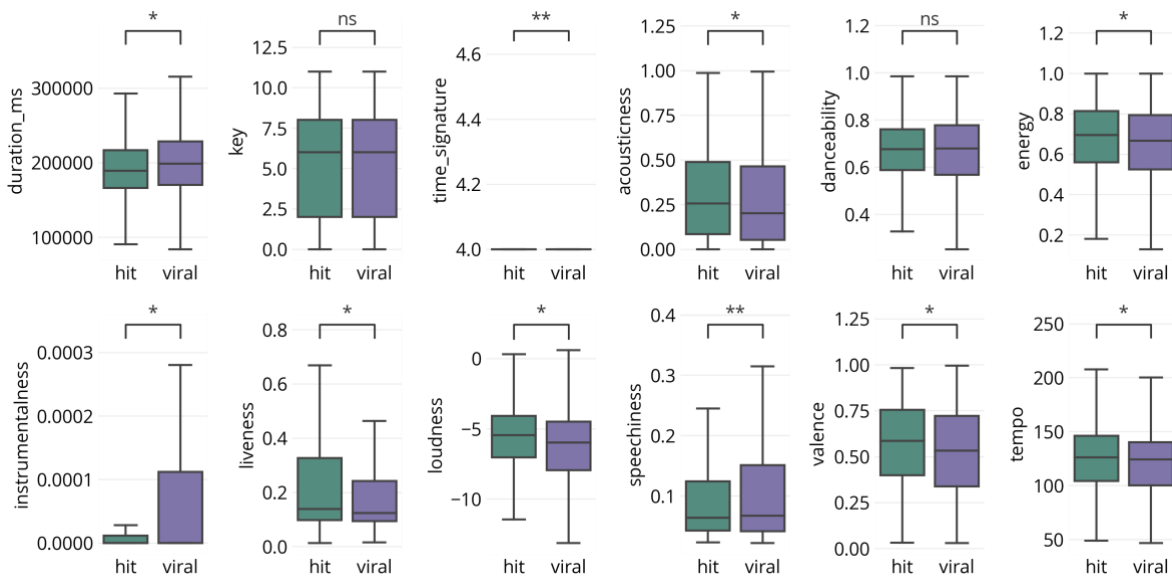
We first verify which songs contain samples from previously released songs, and the ones that are remixes or covers. Table 2 presents the percentage of hit and viral songs associated with pre-existing songs. Overall, most hit or viral songs are original, with only 9.04% and 9.27% of them, respectively, incorporating elements from pre-existing songs. Moreover, there are no substantial differences between hit and viral songs regarding the percentage of remixes and covers as well.

We now look into the music genres of hit and viral songs. Table 3 presents the top 10 most frequent genres of artists who sing both hit and viral songs. We can outline some conclusions. First, in terms of similarities, pop and its variants (*pop nacional* and dance pop) appear in both rankings. This follows a global trend (Oliveira et al., 2020), establishing *pop* artists as the prevailing ones among hit and viral songs. Regional genres (i.e., created in Brazil) also play a key role in the Brazilian market, with a strong presence of *sertanejo* (*sertanejo universitário*, *agronejo*), Brazilian funk (*funk carioca*, *funk rj*) and *arrocha*.¹¹ Indeed, it is common that the Top 10 hit songs in Spotify Brazil are composed entirely (or almost) of songs of artists belonging to these three main genres.

When comparing hits versus viral songs, hip-hop (and its subgenre Brazilian hip-hop) is among the most prominent genres in virals, indicating widespread sharing of songs from artists belonging to this genre on Spotify. For instance, Emicida (one of Brazil's most influential hip-hop artists) boasts 13 viral chart entries but only three hits. Similar patterns are seen in K-pop, with 95 artists in viral charts, contrasting with 16 in hit charts.

¹¹ *Arrocha* is a music genre originated in the Brazilian Northeast that is closely related to *forró* and *axé*.

FIGURE 2 – Boxplots with the distribution of acoustic features' values for hit and viral songs. Significance levels of the Mann-Whitney U test: * for $p < 0.001$; ** for $p < 0.01$; and 'ns' otherwise.



3.2. Acoustic Features

Since the acoustic features provided by Spotify are all represented by numeric values (see Table 1), we analyze the distribution of each one to compare hit and viral songs. Figure 2 illustrates this comparison through boxplots.¹² In addition, we also perform a two-sided Mann-Whitney U test (Mann; Whitney, 1947) to verify the statistical significance of the difference between the distributions for hit and viral songs. In such a test, our null hypothesis is that the distribution underlying the two sets of values is the same. If the p-value of the test is less than a predefined threshold, the null hypothesis is rejected, suggesting a significant difference between the groups.

There are no large differences between hits and virals, but most features present some statistical differences. Such data variance provides valuable insights into the disparities that exist between hit and viral songs, shedding light on the factors that set them apart.

¹² Boxplots visually represent numerical data distribution, skewness, and key summary statistics: minimum score (bottom whisker), lower quartile (25% below the filled area), median (mid-point line), upper quartile (25% above the filled area), and maximum score (upper whisker).

The key findings are as following. Hit songs tend to have a shorter duration on average compared to viral tracks. Indeed, songs have been getting shorter in the last years. Songs with less than three minutes are not uncommon.¹³ For the sake of illustration, “As It Was” by Harry Styles and “Mal Feito - Ao Vivo” by Hugo & Guilherme and Marília Mendonça (i.e., the most listened songs on Spotify globally and in Brazil in 2022) have 2 minutes and 47 seconds, and 2 minutes and 57 seconds, respectively.

On the other hand, hit songs have higher values for acousticness and liveness than viral songs, which means that there are more hits with higher probabilities of being acoustic and being performed live. Such a result may seem counterintuitive at first, as only (or mostly) studio versions are expected to reach the hit status. However, some popular genres in Brazil (such as *sertanejo* and *pagode*) present a specific behavior in which the live versions of songs are the most consumed by listeners. Indeed, the most streamed Brazilian song in 2022 (i.e., “Mal Feito - Ao Vivo” by Hugo & Guilherme and Marília Mendonça) is a live version. This may be a particular characteristic of Brazil, which may not be reflected worldwide. This result corroborates previous work that emphasizes the importance of analyzing other regional markets individually, as each one has its own patterns (Oliveira et al., 2020).

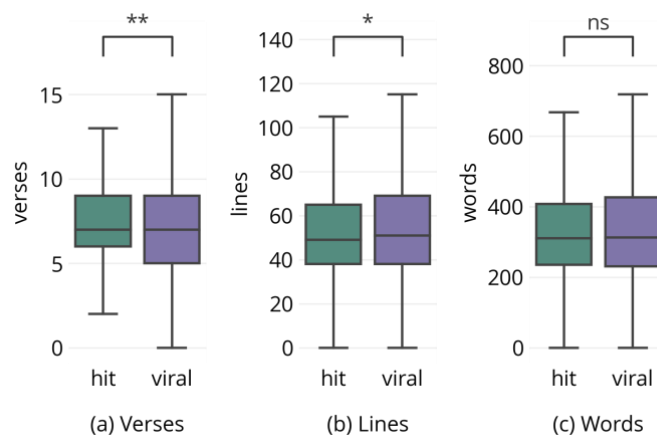
Last, features such as energy, loudness, valance, and tempo reveal that hits are more energetic, more positive, louder, and faster than virals. Such characteristics are often tailored to have broad mass appeal and are more likely to resonate with a large and diverse audience. Viral songs may span a broader range of styles, some of which may not emphasize these qualities as much. Moreover, speechiness values reveal that viral songs have, in general, a higher probability of having spoken words than hits. For instance, chill songs that become a trend in Tik Tok videos and Instagram Reels are widely shared to become viral, but they are not massively streamed to reach the hit charts. An example is the song “Boho Days” of the soundtrack from the movie *tick, tick... BOOM!* (2021), which is currently the third most streamed song from the OST in Spotify.

¹³ Vice (Oct. 2023): <https://bit.ly/3SakvMc>

3.3. Lyrics-related Features

To analyze the songs' lyrics, we consider four dimensions: general features, language, main topics of the lyrics, and their psycholinguistics. Such analyses were performed over the subset of songs with lyrics extracted from Genius (i.e., our reduced dataset, see Section 2.1).

FIGURE 3 – Distribution of the number of (a) verses, (b) lines, and (c) words for hit and viral songs. Significance levels of the Mann-Whitney U test: * for $p < 0.001$; ** for $p < 0.01$; and 'ns' otherwise.



General Characterization. Figure 3 presents the distribution of the number of verses, lines, and words for hit and viral songs. We perform a two-sided Mann-Whitney U test to check whether the difference between the distributions is statistically significant. The results reveal that, although the distribution of verses and words is statistically different, its visual analysis does not show significant differences, i.e., the distributions are similar.

TABLE 4 – Most frequent languages in hit and viral songs in Brazil. The category “Other” includes songs with unknown language.

Hit			Viral		
Language	Songs	%	Language	Songs	%
Portuguese	1,808	47.79%	English	2,673	53.86%
English	1,745	46.13%	Portuguese	1,780	35.87%
Spanish	102	2.70%	Spanish	208	4.19%
Korean	101	2.67%	Korean	201	4.05%
Other	27	0.71%	Other	101	2.04%

Language. Table 4 reveals the main languages of hit and viral songs in Brazil. Portuguese is the most popular language for music hits in Brazil, corroborating the results that show Brazilians tend to consume more local music (Oliveira et al., 2020; Seufitelli et al., 2022). As expected, Brazil is also influenced by external markets. Specifically, English comes as the second most popular language in hit songs. The number of hits in such language is almost the same as the songs in Portuguese, which reveals a very strong influence of other Western countries, mainly the United States. Spanish and Korean are the third and fourth most popular languages, reflecting the growing popularity of Latin genres (mostly due to the proximity to other Latin American countries and their popularization in the US) and K-pop, respectively.

Regarding viral songs, the most frequent languages are the same four languages. However, English is, by far, the prevailing language in viral songs in Brazil. This may reflect the influence of short video platforms (e.g., TikTok and Instagram Reels) on the listening habits in the country (Bastos et al., 2021). A large number of videos posted on such platforms have songs in their background. The more such videos go viral, the more the songs are shared, and people go to Spotify to search and listen to the whole song. A good example of a song that went viral in Brazil following such mechanism is “death bed (coffee for your head)” by Powfu feat. Beabadoobee.

TABLE 5 – Most representative terms (sorted by significance) in the topics inferred by LDA. Offensive terms and swear words are edited. The translation of Portuguese terms and examples of songs for each topic are presented in Appendices A and B, respectively.

	Topic	Songs	English	Portuguese		
			Terms	Topic	Songs	Terms
Hit	H1	64.61%	know, yeah, love, like, never,	H4	20.65%	senta, toma, quer, joga, hoje,
	H2	27.50%	like, yeah, love, want, baby,	H5	24.42%	yeah, tudo, vida, hoje, então,
	H3	7.88%	yeah, baby, love, version,	H6	54.93%	amor, gente, quero, vida,
Viral	V1	20.39%	yeah, like, b*tch, f*ck, sh*t,	V4	22.73%	senta, toma, quer, joga, quero,
	V2	24.16%	yeah, know, love, baby, like,	V5	63.25%	amor, tudo, vida, gente,
	V3	55.45%	know, love, never, like, time,	V6	14.01%	bumbum, então, quer, soca,

Lyrics' Main Topics. Next, we check the lyrics' topics by applying the Latent Dirichlet Allocation (LDA) algorithm,¹⁴ which automatically infers the topics in a set of documents (i.e., the song lyrics). In short, LDA is a probabilistic model that operates by iteratively assigning topics to words in documents and adjusting those assignments based on the observed word-topic and document-topic relationships. Besides the lyrics text, the algorithm also receives a predefined number k of topics as input.

For this analysis, we consider only songs in English and Portuguese, as they represent approximately 90% of hit and viral songs. To extract the topics, we first remove the songs' stop words¹⁵ and annotations (i.e., indications of who sings each part of the song). We then perform the topic coherence metric (Röder; Both; Hinneburg, 2015) to find the best number k of topics. It measures how well-defined and semantically meaningful the identified topics are within a given text (i.e., our lyrics). The higher its value, the higher the coherence and interpretability of topics. Hence, we choose $k = 3$ since it produces the higher values for this metric.

Table 5 summarizes the LDA output for hit and viral songs, with the most representative terms for each topic. In general, both hit and viral English songs present topics related to different facets of romance and love. Terms such as *love* and *baby* appear in four out of seven identified topics. For viral songs, LDA unveils a very specific topic (Topic V1) that contains mostly swear words and other explicit terms. Such a topic prevails in 20.39% of viral songs in English; and such songs are mostly rap and hip-hop, which are genres already known for having more explicit lyrics.

Topics that characterize hit and viral in Portuguese also share similarities, related to romance and love, with the presence of terms such as *amor*, *coração*, *saudade*, *vida* (Topics H5, H6, and V5). Such topics prevail in the dataset, corresponding to 79.35% and 63.25% of hit and viral songs, respectively. However, there is also a relevant set of topics that are related to sexual context (Topics H4, V4, and V6). Similarly to rap and hip-hop in English, such themes are very present in Brazilian funk and, more recently, in *sertanejo universitário* lyrics.

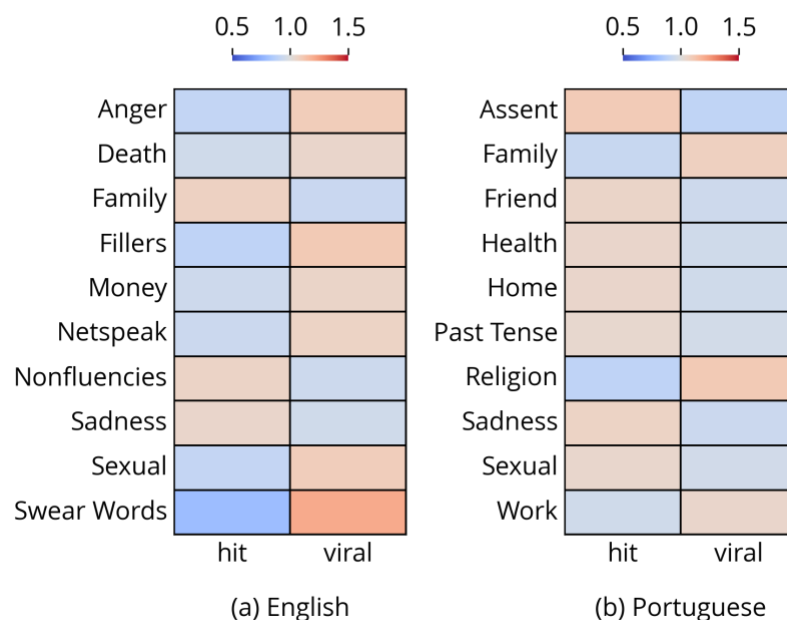
¹⁴ We use the implementation of the Gensim library: <https://radimrehurek.com/gensim/models/ldamodel.html>

¹⁵ Stop words are commonly used words in a language that are filtered out before executing Natural Language Processing (NLP) tasks because they are considered to be of little value for such tasks. Examples include articles, prepositions, and conjunctions.

Psycholinguistic analysis. We now apply the Linguistic Inquiry and Word Count (LIWC) to analyze the psycholinguistic properties of song lyrics to uncover patterns within hit and viral songs. LIWC uses a predefined dictionary of words and linguistic categories to group words and terms within in a given text (i.e., lyrics) into several hierarchical attributes related to linguistic style, affective, and cognitive concepts. We again focus on songs in English and Portuguese.¹⁶

We then identify attributes that characterize both hit and viral songs. To that end, we search for statistical differences across them based on the average frequencies of their respective attributes. Having identified those attributes, we rank them according to their capacity to discriminate across different keywords, estimated by the Gini Coefficient (Yitzhaki, 1979). In such a ranking, we do not consider exclusively linguistic attributes (e.g., linguistic dimensions such as pronouns, auxiliary verbs, and other grammar categories), as our interest lies in emotions and other psychological processes.

FIGURE 4 – Top 10 most discriminative LIWC attributes in viral and hit songs.



¹⁶ For English lyrics, we use LIWC-2015, whereas for Portuguese, we use the version of 2007.

Figure 4 shows a heatmap for the top-10 ranked attributes for (a) English and (b) Portuguese lyrics. The heatmap cells in a column indicate the relative deviation of each attribute for the given keyword from the other keywords. That is, each column (attribute) is normalized following the z-score, i.e., $z = \frac{x - \text{mean}}{\text{std}}$. Thus, each value gets subtracted from the average of the column, then divided by the standard deviation of the column. Therefore, red cells indicate that an attribute is more present in such a category than the average, whereas blue cells mean the opposite. For example, hit songs in English tend to have a higher frequency of swear words compared to the average, whereas viral songs use less such language.

The results show that terms related to family and sadness have a higher frequency in English hit songs compared to virals. In contrast, viral songs often include terms associated with anger, death, money, and swear words. Such themes are commonly found in genres such as rap and hip-hop, which are among Brazil's top viral genres. Moreover, viral English songs have more terms related to sexual content, which again may be related to the influence of viral music genres such as rap and hip-hop.

In Portuguese songs, terms associated with sexual contexts are more frequently found in hit songs than in virals, potentially reflecting the influence of popular music genres. For instance, Brazilian funk and pop songs often use more explicit lyrics in such a context when compared to other genres like *sertanejo* or *arrocha*. However, sexual themes are also present in the latter genres, albeit in a more implicit manner. Viral songs in Portuguese have more terms related to religion, work, and family, which are frequently reported in hip-hop songs.

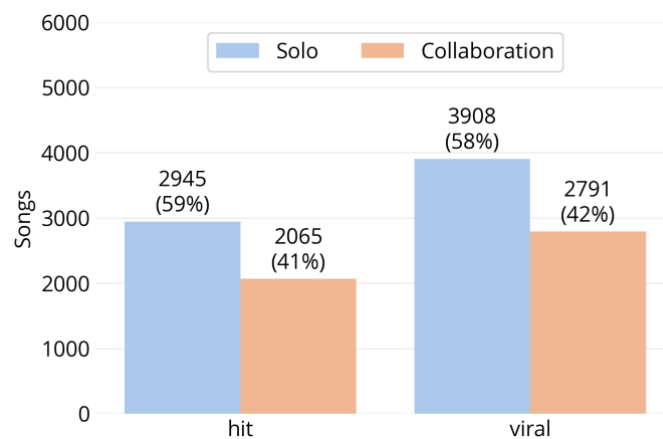
4. Extrinsic Features

We now analyze the songs' extrinsic features: artist-related (Section 4.1) and chart-related ones (Section 4.2).

4.1. Artist-related Features

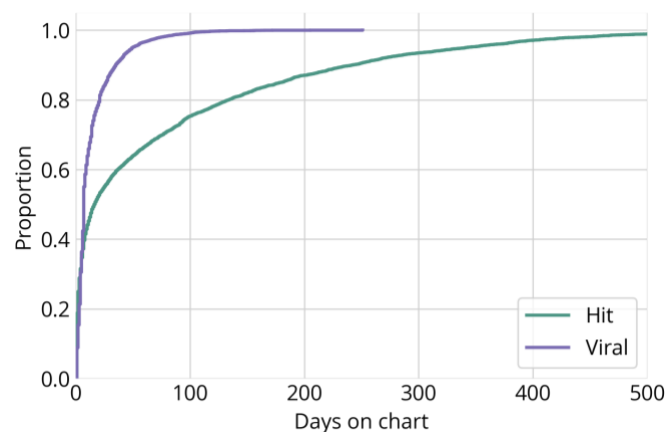
Collaboration has proven to be an important dimension behind musical success (Oliveira et al., 2020; Silva et al., 2022). A song is said to be a collaboration when two or more artists perform it, whether it is a *featuring* or a duet, for example. Conversely, solo songs are sung by only one artist.¹⁷

FIGURE 5 – Proportion of solo and collaboration for hit and viral songs.



In our dataset, Figure 5 shows that the proportion of solos and collaboration within hit and viral songs is similar. The majority of songs are solos, accounting for 59% for hits and 58% for virals. Collaboration occurs in 41% and 42% of hit and viral songs, respectively.

FIGURE 6 – Cumulative Distribution Function of the number of days on charts.

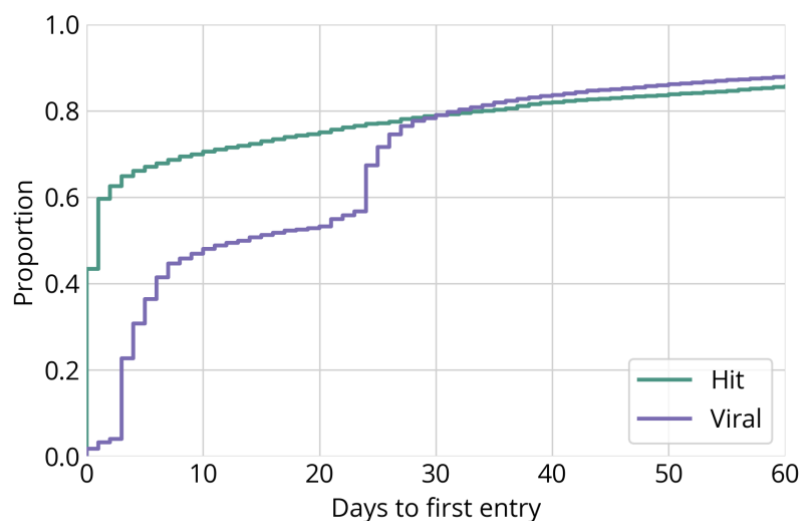


¹⁷ In this work, we consider groups and bands as a single artist.

4.2. Temporal Features

Here, we analyze two specific features to compare hit and viral songs: the number of days a song stays on the charts, and the number of days it takes to reach the charts, also known as *song's debut on the charts*. Regarding the first one, hit songs last much more days on the charts compared to viral ones in Brazil. On average, hit songs stay 75 days (around two months and a half) on the charts, whereas viral songs stay around 14 days (two weeks). The median values are 14 and seven days, respectively. In addition, Figure 6 shows the Cumulative Distribution Function (CDF)¹⁸ of the number of days on charts. The results show that most viral songs last up to 100 days on charts, whereas for hit songs the proportion is around 75%. Such a result confirms the intuition behind what makes viral content: its ephemeral nature (Krijestorac; Garg; Mahajan, 2020). In fact, the popularity of a hit song is much more solid and lasting, whereas viral songs are not.

FIGURE 7 – Cumulative Distribution Function of the number of days from a song's release to its first entry on the charts.



¹⁸ A Cumulative Distribution Function (CDF) is a probability distribution function that describes the probability that a random variable takes on a value less than or equal to a given point. As you move along the x-axis from left to right, the CDF either stays the same or increases, reflecting the cumulative probability of observing a value less than or equal to a specific value of x .

Regarding the days from a song's release to its debut on the charts, the median values for hit and viral songs are one and 14 days, respectively. That is, besides the average being similar, some viral songs take more time to reach the charts. We observe such a pattern in Figure 7, which presents the CDF of the feature over time. Since we are interested in the behavior in the first days after release, we truncate the plot at day 60. For instance, ten days after the release, around 70% of the hit songs had already reached the charts, whereas the value for viral songs is below 60%. The cumulative distributions meet around 30 days after the release.

The position and the duration of a song in the charts can be seen as a measure of its success or virality (but not the only one) (Abel et al., 2010; Shulman; Sharma; Cosley, 2016; Araujo; Cristo; Giusti, 2019). In other words, the longer a song stays in the charts, the more successful/viral it is.

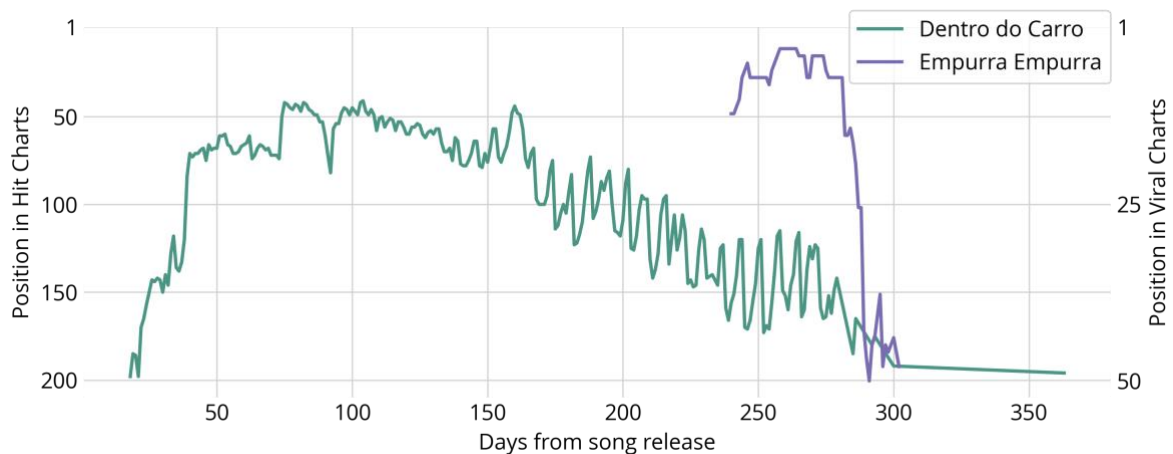
5. Case Studies

To deepen the understanding of the similarities and differences between hit and viral songs, in this section, we add a qualitative analysis of the results by presenting two case studies of songs in the Brazilian market. To do so, we select the songs “Dentro do Carro” by MC Kevin O Chris and “Empurra Empurra” by MC Dricka as instances of hit and viral songs, respectively. Both songs belong to the *funk carioca* genre, one of the main representatives of the Brazilian funk movement. Such a genre is also in the Top 5 most frequent genres in Spotify Brazil for both hits and virals (see Section 3.1).

Regarding acoustic features, the differences are essentially in duration, loudness, and valence. “Empurra Empurra” is longer than “Dentro do Carro” (2min57s versus 4min), being an example of the trend shown that hits have become increasingly shorter. On the other hand, the loudness and valence values are higher in the viral song than in the hit song, contrary to the general trend observed in Section 3.2. The song “Empurra Empurra” is close to the *pancadão* style, a subgenre of Brazilian funk characterized by strong and loud beats, as well as explicit lyrics that tell erotic situations experienced or desired by the singers.

Although both songs have explicit lyrics, there are some relevant differences observed in our analyses. For example, the lyrics of “Dentro do Carro” are considerably shorter than those of “Empurra Empurra” (four and seven verses, respectively). Furthermore, the topic analysis using LDA (Section 3.3) reveals that both songs are predominantly within topics with terms related to sexual content (H4 and V6, respectively). However, when comparing the two topics, V6 presents much more explicit terms than H4. In other words, despite both songs being related to sex, the viral one manages to be much more direct and shameless when talking about the subject.

FIGURE 8 – Position in hit charts for “Dentro do Carro” by MC Kevin O Chris and in viral charts for “Empurra Empurra” by MC Dricka from their releases.



Also, comparing the psycholinguistic processes observed in both songs with LIWC, the viral song presents a greater prevalence of social processes that include terms that suggest human interactions, in addition to negative emotions, mainly anger. This may be due to the more rough and explicit language used in the song. In contrast, the hit song features more terms related to motion, time, and space, as the song title itself refers to a physical space (“Dentro do Carro” translates into “Inside the Car”).

Regarding extrinsic features, the main difference between both songs is in the temporal features, i.e., those related to the behavior of the songs in the charts. Specifically, Figure 8 shows that the hit “Dentro do Carro” enters the charts just 18 days after its release, while “Empurra Empurra” takes 240 days to reach viral status. Furthermore, both songs differ significantly in their permanence

in the charts, corroborating the results of Section 4. Whereas the hit has a solid path of 268 days in the ranking, the viral one is much more ephemeral, staying only 61 days. Moreover, the drop in the chart positions is much more sudden for viral songs than for hits, reinforcing the hypothesis that there is a significant difference in the diffusion processes that occur with hit and viral songs.

6. Discussion

Overall, our results indicate that hit and viral songs exhibit distinct characteristics. While the differences in acoustic features may be subtle, metadata and lyrics-related attributes unveil significant distinctions. Specifically, we find a strong presence of hip-hop and K-pop in viral songs, highlighting the recent trend of popularization of such genres in Brazil. Additionally, psycholinguistic analysis shows that hits in Portuguese have a higher presence of terms related to sexual content; whereas virals are more associated with family, religion, and work.

The existence of many similarities between hits and virals may be partially attributed to the substantial overlap between these two categories. Indeed, there are songs that appear as virals that become hits after a short amount of time. Specifically, there are 1,981 songs that have achieved both hit and viral status, accounting for 39.5% of hit songs and 29.6% of viral songs. Nonetheless, as of the collection date of our dataset, the majority of viral songs had not yet become hit songs.

The temporal features reveal important insights into the behavior of the songs in the charts, as they inform how music is being consumed in Brazil. For the sake of illustration, there is a clear difference in the number of days that viral and hit songs stay in the charts, and also in the period that such songs take to reach them. This suggests the main distinction between hit and viral songs lies mainly in the **diffusion process** itself (i.e., the songs' consumption and their viral spreading), and not necessarily in the intrinsic characteristics of the song. Although this idea may seem intuitive because of the definition of hit and viral, it is essential to support it with concrete data.

7. Conclusion

In this work, we compared hit and viral songs in Brazil by analyzing data obtained from Spotify and enhanced with Genius metadata. While hits and virals share some characteristics, such as their popularity within streaming, we identify specific differences in their intrinsic and extrinsic features. For instance, the analysis of music genres reveals that hit songs are dominated by artists from pop and regional genres such as *sertanejo*, *arrocha*, and Brazilian funk, whereas there is a strong presence of hip-hop and K-pop in viral songs' artists. This is also reflected in other aspects, such as the main themes and emotions present in the lyrics. There are also relevant differences in temporal features, which revealed that virals are much more ephemeral and may take more time to reach the charts than hit songs. We then presented a qualitative analysis with two case studies (one hit and one viral song) that corroborated our findings.

Our results clarify the difference between hit and viral songs as two distinct facets of popularity. In other words, they are two sides of the same coin, with similarities but substantial differences. Indeed, although virality may be a stepping stone to becoming a hit, the majority of viral songs still do not reach the level of hits, highlighting the complexity of the music industry. Once again, this reinforces the definition of virality and success used in this work, as one of the fundamental differences between hit and viral songs lies in the diffusion process itself. As the music industry continues to evolve to follow the changes in technology and audience tastes, this paper offers valuable insights for understanding what drives success on streaming platforms in Brazil and how to shape their success.

Limitations and Future Work. Most limitations of this work are related to the data. In particular, some of the music genres provided by Spotify represent broader artistic movements rather than the musical style itself (i.e., hip-hop). Another limitation lies in the integration of Spotify data with Genius. This may have affected the extent of our analyses, as some songs could not have their lyrics analyzed due to a lack of information. In future work, we plan to improve such integration and perform temporal analyses to verify how hit and viral songs have evolved.

ACKNOWLEDGMENT

This work was supported by Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brazil.

REFERENCES

ABEL, Fabian; DIAZ-AVILES, Ernesto; HENZE, Nicola; KRAUSE, Daniel; SIEHNDEL, Patrick. Analyzing the Blogosphere for Predicting the Success of Music and Movie Products. In: INTERNATIONAL CONFERENCE ON ADVANCES IN SOCIAL NETWORKS ANALYSIS AND MINING (ASONAM), 2010, Odense, Denmark. *Proceedings [...]*. [S. l.]: IEEE, 2010. pp. 276–280. DOI: [10.1109/ASONAM.2010.50](https://doi.org/10.1109/ASONAM.2010.50).

ARAUJO, Carlos Soares; CRISTO, Marco; GIUSTI, Rafael. Predicting Music Popularity on Streaming Platforms. In: SIMPÓSIO BRASILEIRO DE COMPUTAÇÃO MUSICAL (SBCM), 17, 2019, São João del-Rei, Brazil. *Anais [...]*. Porto Alegre: SBC, 2019. pp. 141-148. DOI: [10.5753/sbcm.2019.10436](https://doi.org/10.5753/sbcm.2019.10436).

BASTOS, Hemilly; GIUNTI, Débora Moreira; BENVINDO, Larissa; NASCIMENTO, Alexandre; INOCÊNCIO, Luana. Trends no TikTok e sua influência no streaming musical: os casos Doja Cat e Olivia Rodrigo. In: CONGRESSO BRASILEIRO DE CIÊNCIAS DA COMUNICAÇÃO, 2021, Evento virtual. *Anais [...]*. [S. l.]: INTERCOM, 2021. pp. 1-15.

BISCHOFF, Kerstin; FIRAN, Claudiu S.; GEORGESCU, Mihai; NEJDL, Wolfgang; PAIU, Raluca. Social Knowledge-Driven Music Hit Prediction. In: INTERNATIONAL CONFERENCE ON ADVANCED DATA MINING AND APPLICATIONS (ADMA), 2009, Beijing, China. *Proceedings [...]*. New York: Springer, 2009. pp. 43-54. DOI: [10.1007/978-3-642-03348-3_8](https://doi.org/10.1007/978-3-642-03348-3_8).

BLEI, David M.; NG, Andrew Y.; JORDAN, Michael I. Latent Dirichlet Allocation. *Journal of Machine Learning Research*, [S. l.], v. 3, pp. 993-1022, 2003.

COSIMATO, Alberto; DE PRISCO, Roberto; GUARINO, Alfonso; MALANDRINO, Delfina; LETTIERI, Nicola; SORRENTINO, Giuseppe; ZACCAGNINO, Rocco. The Conundrum of Success in Music: Playing it or Talking About it?. *IEEE Access*, [S. l.], v. 7, pp. 123289-123298, 2019. DOI: [10.1109/ACCESS.2019.2937743](https://doi.org/10.1109/ACCESS.2019.2937743).

DHANARAJ, Ruth; LOGAN, Beth. Automatic Prediction of Hit Songs. In: INTERNATIONAL SOCIETY FOR MUSIC INFORMATION RETRIEVAL CONFERENCE (ISMIR), 2005, London, UK. *Proceedings [...]. [S. l.]: ISMIR, 2005. pp. 488-491.*

DUMAN, Deniz; NETO, Pedro; MAVROLAMPADOS, Anastasios; TOIVIAINEN, Petri; LUCK Geoff. Music we move to: Spotify audio features and reasons for listening. *PLoS ONE, [S. l.], v. 17, n. 9, p. e0275228, 2022. DOI: [10.1371/journal.pone.0275228](https://doi.org/10.1371/journal.pone.0275228)*

GUERINI, Marco; STRAPPARAVA, Carlo; ÖZBAL, Gözde. Exploring Text Virality in Social Networks. In: INTERNATIONAL AAAI CONFERENCE ON WEB AND SOCIAL MEDIA (ICWSM), 5, 2011, Barcelona, Spain. *Proceedings [...]. [S. l.]: The AAAI Press, 2011. pp. 506-509. DOI: [10.1609/icwsm.v5i1.14169](https://doi.org/10.1609/icwsm.v5i1.14169).*

JIANG, Lu; MIAO, Yajie; YANG, Yi; LAN, Zhen-Zhong; HAUPTMANN, Alexander G. Viral Video Style: A Closer Look at Viral Videos on YouTube. In: INTERNATIONAL CONFERENCE ON MULTIMEDIA RETRIEVAL (ICMR), 2014, Glasgow, UK. *Proceedings [...]. New York: ACM, 2014. pp. 193-200. DOI: [10.1145/2578726.2578754](https://doi.org/10.1145/2578726.2578754).*

KONG, Quyu; RIZOIU, Marian-Andrei; WU, Siqi; XIE, Lexing. Will This Video Go Viral: Explaining and Predicting the Popularity of YouTube Videos. In: THE WEB CONFERENCE (WWW), 2018, Lyon, France. *Companion Proceedings [...]. New York: ACM, 2018. pp. 175-178. DOI: [10.1145/3184558.3186972](https://doi.org/10.1145/3184558.3186972).*

KRIJESTORAC, Haris; GARG, Rajiv; MAHAJAN, Vijay. Cross-Platform Spillover Effects in Consumption of Viral Content: A Quasi-Experimental Analysis Using Synthetic Controls. *Information Systems Research, [S. l.], v. 31, n. 2, pp. 449-472, 2020. DOI: [10.1287/isre.2019.0897](https://doi.org/10.1287/isre.2019.0897).*

INTERNATIONAL FEDERATION OF THE PHONOGRAPHIC INDUSTRY. Engaging with music. *[S. l.], 2023. Disponível em: <https://ifpi.org/wp-content/uploads/2023/12/IFPI-Engaging-With-Music-2023_full-report.pdf>. Acesso em: 19 jun. 2024.*

LE COMPTE, Daniel; KLUG, Daniel. "It's Viral!" - A Study of the Behaviors, Practices, and Motivations of TikTok Users and Social Activism. In: ACM CONFERENCE ON COMPUTER SUPPORTED COOPERATIVE WORK AND SOCIAL COMPUTING (CSCW), 2021, Virtual event. *Companion Proceedings [...]. New York: ACM, 2021. pp. 108-111. DOI: [10.1145/3462204.3481741](https://doi.org/10.1145/3462204.3481741).*

LING, Chen; DE CRISTOFARO, Emiliano; STRINGHINI, Gianluca. Slapping Cats, Bopping Heads, and Oreo Shakes: Understanding Indicators of Virality in TikTok Short Videos. In: ACM WEB SCIENCE CONFERENCE (WEBSCI), 2022, Barcelona, Spain. *Proceedings [...]. New York: ACM, 2022. pp. 164-173. DOI: [10.1145/3501247.3531551](https://doi.org/10.1145/3501247.3531551).*

MANN, Henry B.; WHITNEY, Donald R. On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, [S. l.], v. 18, n. 1, pp. 50-60, 1947.

OLIVEIRA, Gabriel P.; SILVA, Mariana O.; SEUFITELLI, Danilo B.; LACERDA, Anisio; MORO, Mirella M. Detecting Collaboration Profiles in Success-based Music Genre Networks. In: INTERNATIONAL SOCIETY FOR MUSIC INFORMATION RETRIEVAL CONFERENCE (ISMIR), 2020, Montreal, Canada. *Proceedings [...]*. [S. l.]: ISMIR, 2020. pp. 726-732.

OLIVEIRA, Gabriel P.; COUTO DA SILVA, Ana Paula; MORO, Mirella M. What makes a viral song? Unraveling music virality factors. In: ACM WEB SCIENCE CONFERENCE (WEBSCI), 2024, Stuttgart, Germany. *Proceedings [...]*. New York: ACM, 2024. pp. 181-190. DOI: [10.1145/3614419.3644011](https://doi.org/10.1145/3614419.3644011).

PRÓ-MÚSICA BRASIL. Mercado Fonográfico Brasileiro 2022. [S. l.], 2023. Disponível em: <https://pro-musicabr.org.br/wp-content/uploads/2023/03/2023-03-20-Mercado-Brasileiros-em-2023.pdf>. Acesso em: 19 jun. 2024.

RÖDER, Michael; BOTH, Andreas; HINNEBURG, Alexander. Exploring the Space of Topic Coherence Measures. In: ACM INTERNATIONAL CONFERENCE ON WEB SEARCH AND DATA MINING (WSDM), 2015, Shanghai, China. *Proceedings [...]*. New York: ACM, 2015. pp. 399-408. DOI: [10.1145/2684822.2685324](https://doi.org/10.1145/2684822.2685324).

SEUFITELLI, Danilo B.; OLIVEIRA, Gabriel, P.; SILVA, Mariana O.; BARBOSA, Gabriel R. G.; MELO, Bruna, C.; BOTELHO, Juliana E.; MELO-GOMES, Luiza; MORO, Mirella M. From Compact Discs to Streaming: A Comparison of Eras within the Brazilian Market. *Revista Vórtex*, [S. l.], v. 10, n. 1, pp. 1-28, 2022. DOI: [10.33871/23179937.2022.10.1.2](https://doi.org/10.33871/23179937.2022.10.1.2).

SEUFITELLI, Danilo B.; OLIVEIRA, Gabriel, P.; SILVA, Mariana O.; MORO, Mirella M. MGD+: An Enhanced Music Genre Dataset with Success-based Networks. In: DATASET SHOWCASE WORKSHOP (DSW), 2023, Belo Horizonte, Brazil. *Anais [...]*. Porto Alegre: SBC, 2023a. pp. 36-47. DOI: [10.5753/dsw.2023.233826](https://doi.org/10.5753/dsw.2023.233826).

SEUFITELLI, Danilo B.; OLIVEIRA, Gabriel, P.; SILVA, Mariana O.; SCOFIELD, Clarise; MORO, Mirella M. Hit song science: a comprehensive survey and research directions. *Journal of New Music Research*, [S. l.], v. 52, n. 1, pp. 41-72, 2023b. DOI: [10.1080/09298215.2023.2282999](https://doi.org/10.1080/09298215.2023.2282999).
SHULMAN, Benjamin; SHARMA, Amit; COSLEY, Dan. Predictability of Popularity: Gaps between Prediction and Understanding. In: INTERNATIONAL AAAI CONFERENCE ON WEB AND SOCIAL MEDIA (ICWSM), 10, 2016, Cologne, Germany. *Proceedings [...]*. [S. l.]: The AAAI Press, 2016. pp. 348-357. DOI: [10.1609/icwsm.v10i1.14748](https://doi.org/10.1609/icwsm.v10i1.14748).

SILVA, Mariana O.; OLIVEIRA, Gabriel P.; SEUFITELLI, Danilo B.; LACERDA, Anisio; MORO, Mirella M. Collaboration as a Driving Factor for Hit Song Classification. In: BRAZILIAN SYMPOSIUM ON MULTIMEDIA AND WEB (WEBMEDIA), 2022, Curitiba, Brazil. *Anais [...]*. New York: ACM, 2022. pp. 66-74. DOI: [10.1145/3539637.3556993](https://doi.org/10.1145/3539637.3556993).

SILVA, Mariana O.; OLIVEIRA, Gabriel P.; SEUFITELLI, Danilo B.; MORO, Mirella M. Temporal Success Analyses in Music Collaboration Networks: Brazilian and Global Scenarios. *Revista Vórtex, [S. l.]*, v. 11, n. 2, pp. 1-27, 2023. DOI: [10.33871/23179937.2023.11.2.7185](https://doi.org/10.33871/23179937.2023.11.2.7185).

TAUSCZIK, Yla R.; PENNEBAKER, James W. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology, [S. l.]*, v. 29, n. 1, pp. 24-54, 2010. DOI: [10.1177/0261927X09351676](https://doi.org/10.1177/0261927X09351676).

TSIARA, Eleana; TJORTJIS, Christos. Using Twitter to Predict Chart Position for Songs. In: IFIP INTERNATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE APPLICATIONS & INNOVATIONS (AIAI), 2020, Neos Marmaras, Greece. *Proceedings [...]*. New York: Springer, 2020. pp. 62-72. DOI: [10.1007/978-3-030-49161-1_6](https://doi.org/10.1007/978-3-030-49161-1_6).

VAZ DE MELO, Gabriel B.; MACHADO, Ana F.; CARVALHO, Lucas R. de. Music consumption in Brazil: an analysis of streaming reproductions. *PragMATIZES - Revista Latino-Americana de Estudos em Cultura, [S. l.]*, v. 10, n. 19, pp. 141-169, 2020. DOI: [10.22409/pragmatizes.v10i19.40565](https://doi.org/10.22409/pragmatizes.v10i19.40565).

YITZHAKI, Shlomo. Relative Deprivation and the Gini Coefficient. *The Quarterly Journal of Economics, [S. l.]*, v. 93, n. 2, pp. 321-324, 1979. DOI: [10.2307/1883197](https://doi.org/10.2307/1883197).

ABOUT THE AUTHORS

Gabriel P. Oliveira is a Ph.D. student in the Computer Science Graduate Program at Universidade Federal de Minas Gerais, Brazil. He received the B.Sc. and M.Sc. degrees in Computer Science from UFMG in 2018 and 2021, respectively. He is currently a member of the Laboratory of Interdisciplinary Computer Science (CS+X) and Data Scientist in the Analytical Capabilities Project. His research interests include social computing, data science and social network analysis, with a strong emphasis on collaborative domains. ORCID: <https://orcid.org/0000-0002-7210-6408>. E-mail: gabrielpoliveira@dcc.ufmg.br

Ana Paula Couto da Silva earned her PhD degree in Computer and System Engineering from the Federal University of Rio de Janeiro (2006). She was with IRISA-Rennes in 2005 and 2007, and with Politecnico di Torino in 2008 (as post-doc), 2016 (as visiting professor) and 2018 (as visiting researcher). She is with the Computer Science Department at the

Federal University of Minas Gerais since 2013 as associate professor. She has productivity in research scholarship (Bolsa PQ - CNPq) level 2. Her areas of interest are in the field of network science, social computing and analysis of computer systems and green networking. ORCID: <https://orcid.org/0000-0001-5951-3562>. E-mail: ana.coutosilva@dcc.ufmg.br

Mirella M. Moro, PhD, works in the Computer Science Department at Universidade Federal de Minas Gerais, Brazil. Her research interests include data-driven research, computing on culture, gender diversity, and education in Computer Science. She is also an advocate for increasing women participation in Computing and Technology and coordinates the Digital Girls Program at the Brazilian Computer Society. ORCID: <https://orcid.org/00000-0002-0545-2001>. E-mail: mirella@dcc.ufmg.br

CREDIT TAXONOMY

1st Author			
X	Conceptualization		Resources
X	Data curation	X	Software
X	Formal Analysis		Supervision
	Funding acquisition	X	Validation
X	Investigation	X	Visualization
X	Methodology	X	Writing – original draft
	Project administration		Writing – review & editing

2nd Author			
X	Conceptualization	X	Resources
	Data curation		Software
X	Formal Analysis	X	Supervision
X	Funding acquisition	X	Validation
	Investigation		Visualization
X	Methodology		Writing – original draft
X	Project administration	X	Writing – review & editing

3rd Author			
X	Conceptualization	X	Resources
	Data curation		Software
X	Formal Analysis	X	Supervision
X	Funding acquisition	X	Validation
	Investigation		Visualization
X	Methodology		Writing – original draft
X	Project administration	X	Writing – review & editing

APPENDICES

A. Translation of Portuguese terms

Here, we present the translation of Portuguese terms of the topic analysis of Table 5. Such terms carry several meanings that are directly linked to the Brazilian culture, which may not be translated accurately into English.

H4. sit, take, want, play, today, floor, butt, butt, then, go down

H5. yeah, everything, life, today, then, here, God, want, love, forever

H6. love, people, want, life, heart, everything, to miss, nothing, time, mouth

V4. sit, take, want, play, want, go down, yeah, today, can, call

V5. love, everything, life, people, want, today, here, time, because, nothing

V6. butt, then, want, punch, everything, deny, mouth, floor, four, put

B. Examples of songs for each LDA topic

H1. the light is coming (feat. Nicki Minaj) by Ariana Grande

H2. Walk It Talk It (feat. Drake) by Migos

H3. What Lovers Do (feat. SZA) by Maroon 5

H4. Dentro do Carro by MC Kevin O Chris

H5. Melhor Dia 6 - Destino (part. Jovem Dex, Yunk Vino, The Boy, Wall Hein e Clovis Pinho) by Marcos Baroni

H6. Assim Nasce Um Bêbado by Luan Santana

V1. No Stylist (feat. Drake) by French Montana

V2. All The Stars (feat. SZA) by Kendrick Lamar

V3. Signs of Life by Arcade Fire

V4. Trava na Pose, Chama no Zoom, Dá um Close (feat. Mc Rennan) by DJ Patrick Muniz, Dj Olliver, Mc Topre

V5. Canção Infantil by Cesar Mc

V6. Empurra Empurra by MC Dricka