Modeling music popularity as an epidemic: insights from the Brazilian market

Modeling music popularity as an epidemic: insights from the Brazilian market

Gabriel P. Oliveira[1], **Luca Vassio**[2], **Ana Paula Couto da Silva**[1], **Mirella M. Moro**[1]

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

[2]Politecnico di Torino – Turin, Italy

`gabrielpoliveira@dcc.ufmg.br, luca.vassio@polito.it, {ana.coutosilva,mirella}@dcc.ufmg.br`

Abstract. *Social networks have drastically changed the music market scenario, enabling songs to reach massive audiences in record time. Given the fast-paced nature of music popularity, this work investigates whether epidemic models can effectively capture how songs gain traction. We apply the Susceptible-Infected-Recovered (SIR) model to analyze music virality and success in Brazil. Virality reflects a song's rapid surge in popularity, whereas success represents its long-term endurance. By comparing the model's fit for both trajectories, we assess its strengths and limitations in capturing music popularity trends. Our findings reveal that SIR provides a better fit for virality than for long-term success, reinforcing the differences between both processes and, consequently, offering a new perspective on how songs become popular in the digital age.*

1. Introduction

The growing influence of social networks has fundamentally reshaped the way information, trends, and cultural phenomena spread in society. Platforms such as TikTok and Instagram enable content to reach massive audiences in a short amount of time, often without the involvement of traditional media or industry [Compte and Klug 2021, Ling et al. 2022]. This fast and decentralized mode of dissemination has particularly impacted the music industry, where a song can go from complete obscurity to mainstream success within days, driven by user-generated content and algorithmic recommendations.

Music consumption has evolved significantly over time, moving from physical formats such as vinyl records and CDs to digital downloads and, more recently, streaming services. A recent report from the International Federation of the Phonographic Industry (IFPI) reveals that 73% of people listen to music using audio streaming services.[1] Such platforms not only offer instant access to a vast catalog of songs but also provide real-time data on listening trends. This shift has made music popularity much more dynamic, allowing new hits to emerge and old songs to become popular again through viral trends.

A notable example in Brazil is the song "Batom de Cereja" by Israel & Rodolffo, which became a massive hit in 2021. The song gained traction after being frequently played on the reality show *Big Brother Brasil*, leading to a surge in TikTok videos featuring its melody. This viral moment was reflected in its following success, as it later

[1]IFPI Engaging with Music: https://www.ifpi.org/wp-content/uploads/2023/12/IFPI-Engaging-With-Music-2023_full-report.pdf

became the most-listened song on streaming platforms.[2] This example illustrates how online engagement can fuel a song's rapid rise, besides traditional industry promotion.

The dynamics of content spreading on social platforms present several similarities to the dissemination of infectious diseases. In fact, this has been studied in related contexts, including song downloads [Nika et al. 2015, Rosati et al. 2021] and video views [Sachak-Patwa et al. 2018]. Indeed, epidemic models may also be a promising framework for analyzing music popularity since they describe how something (i.e., whether a virus, an idea, or a song) passes through a population, making them well-suited for studying trends that rise and fall over time. Therefore, in this work, we investigate the dynamics of music popularity on streaming platforms by answering the following question: *"Can music popularity in Brazil be effectively modeled as an epidemic process?"*

To address this question, we apply a Susceptible-Infected-Recovered (SIR) model to the virality and success trajectories of songs in the Brazilian market. We consider both concepts as distinct yet interconnected facets of music popularity. Virality is related to a song's rapid and explosive spread, typically driven by short-term social media trends, memes, or challenges [Guerini et al. 2011]. In contrast, success is a more sustained form of popularity, characterized mainly by streaming numbers, album sales, or radio airplay [Seufitelli et al. 2023a]. While both viral and hit songs can achieve widespread recognition, their trajectories differ significantly, requiring distinct approaches to their analysis.

The rest of this paper is organized as follows: Section 2 reviews related work, whereas Section 3 details our methodology. Section 4 presents the results, comparing virality and success. Section 5 discusses such findings and some limitations of epidemic modeling in this context. Finally, Section 6 concludes with future research directions.

2. Related Work

Understanding the dynamics and the factors that lead to music popularity has been an emerging topic in research, resulting in a brand new research field called Hit Song Science (HSS). The work of [Dhanaraj and Logan 2005] was the first to investigate how song-related features contributed to their success. Since then, several perspectives have been used to understand musical popularity, including chart performance and economic and engagement metrics [Seufitelli et al. 2023a]. More recently, with the emergence of social networks, music virality has also become relevant, being associated with other different concepts such as word-of-mouth [Sharma et al. 2011], spreading [Cha et al. 2012, Rosati et al. 2021], and viral marketing [Barbieri and Bonchi 2014].

Different music markets exhibit unique characteristics, each one with its own behavior and listening patterns [Mondelli et al. 2018, Pereira et al. 2018]. The Brazilian market is no exception, and even different regions within the country have their own particular music tastes [Moura et al. 2024]. Like many countries, Brazil transitioned from physical to digital music consumption in the 2000s and 2010s, a shift that, along with social media, contributed to the rise of new genres and consuming patterns [Barbosa et al. 2021]. Furthermore, recent work on the Brazilian music market used data from streaming platforms to provide evidence for the difference between music virality

[2] Correio Braziliense: https://www.correiobraziliense.com.br/diversao-e-arte/2021/07/4936571-batom-de-cereja-de-israel---rodolffo-e-maior-hit-do-1-semestre.html

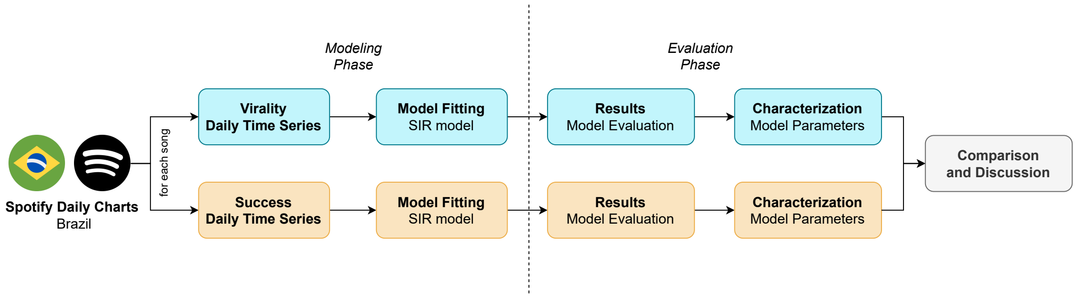

Figure 1. Methodology used for representing music popularity (both success and virality) as an epidemic process.

and success as distinct facets of popularity [Oliveira et al. 2024a], as well as the existence of a temporal relationship between them [Oliveira et al. 2024b].

Besides understanding the differences between music virality and success, it is also necessary to understand in depth how such processes happen online. Following the work of [Centola and Macy 2007], behavior diffusion online has been addressed either as simple or complex contagion processes. In simple contagion, a single exposure is sufficient for an individual to adopt a behavior, whereas complex contagion requires multiple exposures or reinforcement from different sources to trigger adoption. In this context, recent works have used complex models to represent several online phenomena, including the spread of information [Mønsted et al. 2017] and political symbols [Medina 2023].

Although complex approaches have also been used to understand culture-related phenomena [Ternovski and Yasseri 2020], most of the existing research uses simple epidemic models to understand the spread of ideas and trends online. In particular, the Susceptible-Infected-Recovered (SIR) model has been used to analyze music diffusion on social media [Nika et al. 2015, Rosati et al. 2021, Li and Shao 2024], since it captures contagion rates and peak popularity periods. More complex models have also been explored to refine this understanding [Sachak-Patwa et al. 2018]. Whereas such approaches have successfully used metrics such as song downloads and video views, their applicability to the current streaming environment remains relatively unexplored.

Overall, although there is a considerable amount of research on music virality, success, and the application of epidemic models in different contexts, no previous study has focused explicitly on song popularity in the streaming era. While previous works have explored music dissemination and online trends, they have yet to apply epidemic models to data from streaming platforms. Studying the Brazilian market is also relevant, not only because it is one of the largest music markets in the world, but also because it presents its peculiar behavior and trends. Therefore, this work represents a step forward in assessing the effectiveness of an epidemic model in capturing music popularity in Brazil, also by treating virality and success as distinct but interconnected aspects of such a phenomenon.

3. Methodology

Our methodology is illustrated by Figure 1, divided into two phases: modeling and evaluation. This section focuses on the modeling phase, covering the dataset (Section 3.1), time series construction (Section 3.2), and the epidemic model (Section 3.3). The evaluation and results are discussed in the following section.

3.1. Data

With the consolidation of the Web and social platforms as one of the most important means of content dissemination, streaming services have been playing an essential role in the consumption and spreading of music. According to the International Federation of the Phonographic Industry, in 2023, 67.9% of the global recorded music revenues came from streaming (considering ad-supported and subscriptions).[3] In Brazil, this percentage is even higher, reaching 86.2% of the revenue in 2022.[4] Therefore, given the current scenario, it is reasonable to consider streaming services as data sources for user behavior and music consumption at both Global and regional levels.

In this work, we consider data from Spotify, one of the most popular streaming services in the world, with over 675 million users (including 263 million subscribers) in more than 180 markets.[5] For each market, the platform produces daily rankings for the most **successful** and **viral** songs (i.e., Top 200 and Viral 50, respectively). The methodology for building the rankings is different, reinforcing the difference in the concepts of success and virality discussed in Section 1. Whereas the Top 200 is basically the ranking of the most streamed songs, the Viral 50 captures songs gaining attention on the platform, and is calculated considering factors such as the rise in shares, plays, and the number of people who have recently discovered the songs.[6]

Specifically, we obtain data from the Music Genre Dataset (MGD+), a curated dataset based on enhanced data from Spotify Charts [Seufitelli et al. 2023b]. It contains daily charts for each music market from January 2017 to March 2022. Although Spotify continues to produce rankings, it is not possible to update the dataset with more recent data due to significant changes on the Spotify Charts platform that do not allow downloading the charts. We consider as hit (i.e., successful) songs all the songs within the Top 200 charts, whereas virals are those in the Viral 50 charts, regardless of the position.

Overall, we evaluate 1,895 daily success and viral charts in Brazil, comprising 9,728 distinct songs from 5,126 artists. Furthermore, since we want to compare our epidemic methodology applied to success and virality, we reduce the final set of songs to the 1,977 songs present in both charts, i.e., songs that are both hit and viral.

3.2. Time Series Modeling

We now build the time series from the set of viral and hit songs that represent their virality and success, respectively. We use the same approach proposed by [Oliveira et al. 2024b], who use a song's chart position to indicate its popularity. For each song, its time series starts on its release date or the first chart considered (i.e., January 1st, 2017) in case it is older. Similarly, the last date for all songs' time series is the last considered chart. Each point represents the daily performance in the Top 200 and Viral 50 charts.

The viral/success performance is measured using the rank score, a metric calculated solely by a song's position on the charts. For a song on a position i, its value is

[3] IFPI Global Music Report: https://globalmusicreport.ifpi.org/

[4] Pró-Música Brasil: https://pro-musicabr.org.br/wp-content/uploads/2023/03/2023-03-20-Mercado-Brasileiros-em-2023.pdf

[5] Spotify (March 2025): https://newsroom.spotify.com/company-info/

[6] Although providing such a description, Spotify does not inform the exact formula used for ranking the viral songs, which can limit our analyses and interpretations. This limitation is further discussed in Section 6. For more information on chart making, see: https://support.spotify.com/us/artists/article/charts/

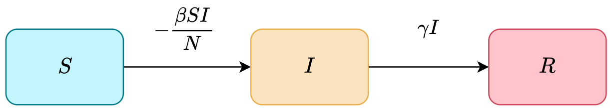

Figure 2. Overview of the SIR model.

simply given by $rank\_score(i) = max\_rank - i + 1$, where max_rank represents the lowest possible position (200 for the success chart and 50 for the viral one). For instance, a song that reaches the #5 spot on the Viral 50 chart (i.e., the success chart) obtains a rank score of 46, as $rank\_score(5) = 50 - 5 + 1$. If the song is not present in the charts on a specific date, we set its rank score to zero, regardless of the actual value of success/virality (that we cannot measure). We employ this metric to ensure that songs that achieve top positions receive the highest rank scores, while those ranked lower receive progressively smaller values. In other words, using a rank score allows a straightforward comparison of chart performance based solely on numerical scores.

3.3. SIR Model

Having defined how to measure the temporal evolution of music popularity (i.e., virality or success), the next step is to represent it as a dissemination process. In this work, we consider epidemiological models, which have been extensively used to represent not only the spread of diseases [Cançado et al. 2024], but also the dissemination of human behavior and information on social platforms [Raponi et al. 2022]. Our assumption is that music popularity is closely related to how people consume it, and once a person listens to a song and gets impacted by it (either positively or negatively), they are more prone to share – or at least talk about – it with others.

Here, we use the SIR model, the simplest model for the spread of infections (Figure 2). According to the definition used by [Bjørnstad et al. 2020], this model considers a three-state epidemic, in which individuals can be either susceptible (S), infected (I) or recovered (R). This model has no probabilistic component and considers a closed epidemic (i.e., without births and deaths), with a fixed population of $N = S + I + R$. The number of people in each state is a function of time t, and individuals in the susceptible state can be infected at a rate β, as the infected ones recover at a rate γ. The rates of change of the susceptible, infected and recovered states are given by Equations 1, 2, and 3, respectively.

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -\frac{\beta SI}{N} \quad (1) \qquad \frac{\mathrm{d}I}{\mathrm{d}t} = \frac{\beta SI}{N} - \gamma I \quad (2) \qquad \frac{\mathrm{d}R}{\mathrm{d}t} = \gamma I \quad (3)$$

Besides its simplicity, the SIR model has been used in similar contexts [Rosati et al. 2021, Li and Shao 2024], making it suitable also for this task. In addition, using such a model allows for interpreting its states and parameters with actual meanings. For instance, for each song:

- **Susceptible** (S) individuals are those who have not been exposed to the song yet, which can happen through recommendation, playlists, and social networks;
- **Infected** (I) individuals are the ones who are actually spreading the song's popularity by streaming (in the case of success) or sharing (for virality) it; and

- **Recovered** (R) individuals are people who had already listened to the song but lost interest and stopped consuming it.

Furthermore, the rate β (infection rate) can be interpreted as the song's adoption rate, which is the likelihood of an individual starting to consume and/or sharing it. Similarly, γ (recovery rate) may represent the song's forgetting rate, which is the chance of an individual stopping streaming or sharing it.

4. Evaluation and Results

This section presents the second phase of the methodology of Figure 1, which comprises the model evaluation and a discussion on the model parameters after fitting. After detailing the experimental setup (Section 4.1), we evaluate the SIR fitting results regarding the original time series data (Section 4.2). Then, we analyze in depth the model parameters for the songs by characterizing their distribution and also discussing the meaning of such values in the music dissemination context (Section 4.3).

4.1. Experimental Setup

We perform some preprocessing steps to prepare the data to serve as input for the SIR model. First, we apply a 7-day moving average to smooth fluctuations in the time series. Next, we perform min-max normalization, scaling values to $[0, 0.5]$ to ensure comparability across the success and viral rankings. We set the maximum value to 0.5 instead of 1.0 because the time series represents the number of infected individuals, and it would be unrealistic to assume that the entire population becomes infected. In addition, on days when the popularity value is zero (i.e., the song did not appear on the charts), we introduce a small amount of noise by setting the value to 0.001. We do so because a song's absence from the charts does not necessarily mean it had no popularity that day.

Next, we define the initial conditions necessary for fitting the SIR model to each time series. Specifically, we set the total population N to 1.0 (since the values are normalized) and the initial infected people I_0 to the first observed value in the time series. The initial susceptible population S_0 is set as $N - I_0$, and the initial recovered count R_0 is assumed to be zero at the start.

Regarding the implementation, we use the *SciPy* Python library[7] to estimate the model parameters [Virtanen et al. 2020]. We use the `integrate.odeint` function for numerical integration of the differential equations and `optimize.curve_fit` for parameter fitting, employing the least squares method with a defined inferior bound of 0 and an initial guess of 0.5 for each SIR parameter (i.e., β and γ). The model's accuracy is evaluated using the Root Mean Squared Error (RMSE) to quantify the deviation between the real observed data and the fitting result, ranging from 0 (perfect fit) to 0.5 (worst case).

4.2. Model Evaluation

We now evaluate how well the SIR model fits both virality and success curves for our songs. Besides checking whether this model is suitable for representing music popularity in general, we aim to discover which of the two processes (i.e., virality and success) is better described by such a model. The average RMSE is 0.028 ± 0.001 for virality and

[7]SciPy: https://scipy.org/

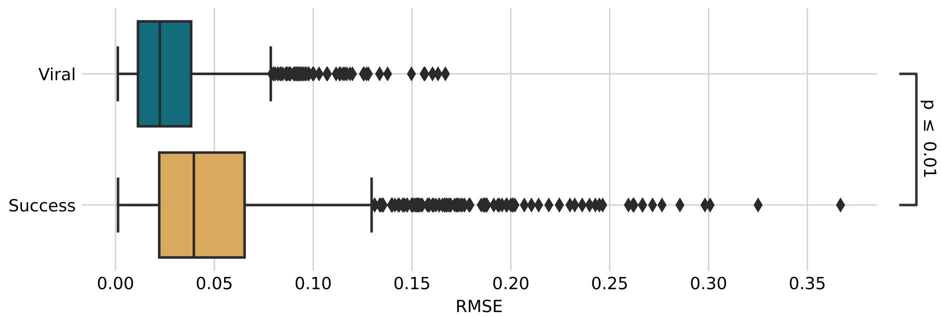

Figure 3. RMSE for virality and success curves using the SIR model. Significance is calculated using the Mann-Whitney U test (p-value = 1.845e-64).

Table 1. Top 5 songs with highest RMSE values for the SIR fitting on their virality and success time series.

	Virality			
Song	**Artists**	β	γ	**RMSE**
Vem Me Satisfazer	MC Ingryd, DJ Henrique da VK	0.532	0.492	0.167
Ilusão	C4000, Brocasito	0.470	0.450	0.163
Anota Placa	Tz da Coronel, DJ PurpleRain.REC, Vinta	0.492	0.471	0.160
Life Goes On	Oliver Tree	0.486	0.450	0.156
Pedaço de Mim	OIK, BIN, IssoQueÉSomDeRap	0.403	0.114	0.156
	Success			
Song	**Artists**	β	γ	**RMSE**
Água Nos Zói - Ao Vivo	Clayton & Romário, Jorge & Mateus	0.805	0.225	0.367
abcdefu	GAYLE	0.514	0.478	0.325
Investe Em Mim	Jonas Esticado	0.481	0.470	0.301
Idiota	Jão	0.468	0.350	0.298
Louis V, Menina Linda	Sidoka, Intactoz Corp.	0.510	0.486	0.285

0.052 ± 0.002 for success, both with 95% confidence interval. This statistically significant difference suggests that the error is higher for success than for virality, indicating that SIR models better capture the dynamics of virality compared to success.

This result is further supported by the analysis of the RMSE distribution, as shown in Figure 3. To assess the statistical significance of the difference between the error distributions for virality and success, we perform a two-sided Mann-Whitney U test [Mann and Whitney 1947]. This non-parametric test is suitable for comparing two independent samples without assuming a normal distribution. In this case, the null hypothesis states that the two distributions come from the same underlying population, meaning they have no significant difference. The results show that such a hypothesis is rejected with $p \leq 10^{-5}$, confirming that the error distribution differs significantly between them and reinforcing that SIR models are better suited for modeling virality than long-term success.

To put it in more practical terms, Table 1 lists the five songs with the highest RMSE values for the SIR model fitting in both virality and success time series. In fact, the success time series present higher errors compared to the virality ones, reinforcing the idea that the SIR model is better suited to represent short-term popularity trends. For

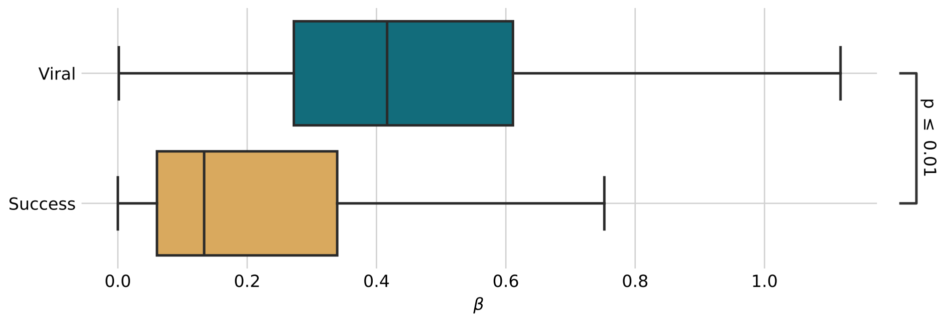

Figure 4. **β values for virality and success curves using the SIR model. Significance is calculated using the Mann-Whitney U test (p-value = 3.875e-139). Outliers are omitted for visual clarity (viral has 137 outliers, maximum is 3.155; success has 118 outliers, maximum is 10.716).**

example, the error for the success fit for *Água nos Zói - Ao Vivo* is more than two times higher than the error for the virality of *Vem Me Satisfazer*. Although this may be partially because viral time series have fewer points different from zero, it also suggests that the SIR model may not capture the dynamics of long-term success well.

4.3. Model Parameters

One of the most significant advantages of using epidemic models such as SIR is the ability to interpret their parameters in a meaningful way, providing insights into the dynamics of music consumption. Specifically, in this section we discuss the results of the two parameters of the model (i.e., β and γ) and two other relevant variables derived from them: the infectious period and the basic reproduction number.

Infection rate (β). In the context of music consumption, it quantifies the speed at which a song gains new listeners, whether in terms of virality or success. A high β in the viral context suggests that a song spreads rapidly through user sharing, quickly capturing public attention. Conversely, in the success context, a high β indicates that the song is steadily attracting a broad audience and streams, possibly due to sustained marketing efforts or inclusion in major playlists. While a high infection rate reflects strong adoption, it does not necessarily guarantee long-term retention, as the song's longevity also depends on the balance between infection (β) and recovery (γ) rates.

Figure 4 shows the distribution of the β values for both virality and success time series. The median values for this parameter are 0.416 for viral and 0.133 for success. This indicates that, in general, viral songs spread more rapidly compared to hit songs, reflecting the nature of viral phenomena, which tend to experience a sharp rise in popularity driven by social media trends and organic sharing. In contrast, the lower β for success suggests a more gradual adoption process, where songs build an audience over time, often influenced by sustained marketing efforts and playlist placements.

Recovery rate (γ). It quantifies the speed at which people stop streaming or sharing a song, either because they lose interest or move on to new trends. A high γ value for virality suggests that viral songs tend to fade as quickly as they rise, reflecting short-lived trends where engagement is intense but temporary. On the other hand, a high γ

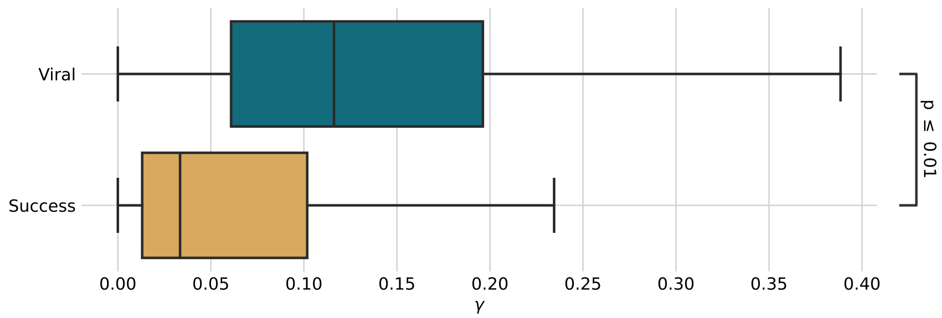

Figure 5. **γ values for virality and success curves using the SIR model. Significance is calculated using the Mann-Whitney U test (p-value $=$ 1.435e-104). Outliers are omitted for visual clarity (viral has 65 outliers, maximum is 1.780; success has 231 outliers, maximum is 5.359).**

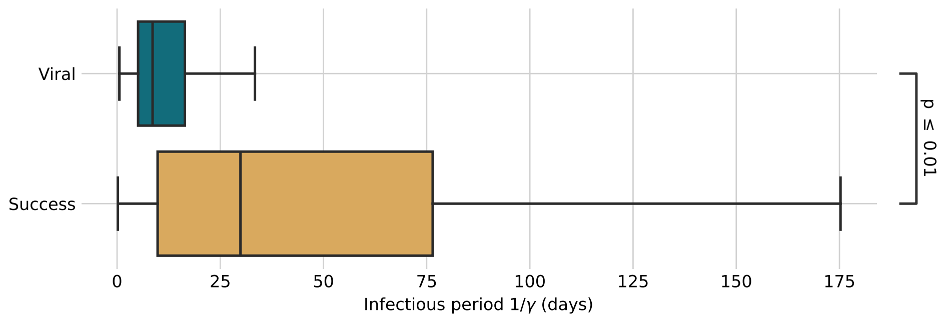

Figure 6. **Infectious period for virality and success curves using the SIR model. Significance is calculated using the Mann-Whitney U test (p-value $=$ 1.435e-104). Outliers are omitted for visual clarity (viral has 104 outliers, maximum is 1.313e+20; success has 57 outliers, maximum is 3.047e+17).**

value for success indicates that even popular songs in the Top 200 ranking can experience fast declines, possibly due to competition from new releases or a loss of promotional momentum. Conversely, a low value for this parameter implies longer-lasting popularity, with the song maintaining steady engagement over time.

The distribution for this parameter is illustrated by Figure 5. The average γ values are 0.116 for virality and 0.033 for success, meaning that, on average, viral songs tend to lose listeners more quickly than successful songs, reinforcing the idea that viral popularity is often short-lived. The higher values for virality suggest that such songs experience a rapid rise and fall in engagement, whereas the lower values for success imply that hit songs in the Top 200 generally sustain their popularity for longer periods, benefiting from continued audience engagement and marketing efforts.

Infectious period $(1/\gamma)$**.** From the two parameters directly obtained from the SIR model, it is also possible to derive other variables with relevant meaning in the music popularity context. For example, $1/\gamma$ informs the infectious period of an individual, i.e., the number of days in which they are impacted by a song (whether by sharing it or streaming it). Therefore, the lower the γ value, the longer an individual consumes it. Conversely, a

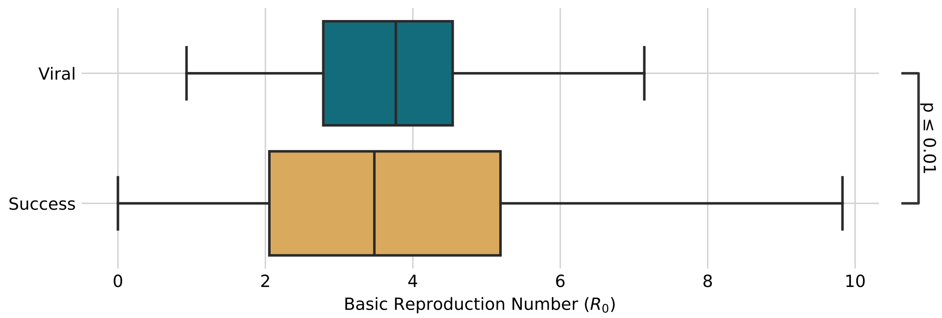

Figure 7. R_0 **for virality and success curves using the SIR model. Significance is calculated using the Mann-Whitney U test (p-value = 5.626e-03). Outliers are omitted for visual clarity (viral has 82 outliers, maximum is 3.699e+19; success has 167 outliers, maximum is 1.941e+16).**

higher γ indicates a shorter infectious period, meaning that listeners lose interest more quickly, characterizing songs that experience a rapid but fleeting surge in popularity.

The distribution of the values is presented in Figure 6. The median value for the infection period is around 9 days for viral songs and 30 days for hits. Again, such a finding reinforces the fact that viral songs experience a rapid but short-lived surge in popularity, with listeners quickly moving on to new trends. In contrast, hit songs tend to sustain listener interest for a much longer duration. The contrast between such values and distributions reinforces the idea that virality and long-term success follow distinct popularity trajectories within the music industry.

Basic reproduction number (R_0). This number is calculated as the ratio β/γ and quantifies a song's potential to spread organically. In other words, a high R_0 for virality suggests that a viral song has a strong capacity to be rapidly shared and adopted, which often may be driven by social media trends, challenges, or word-of-mouth diffusion. Conversely, a high R_0 for success indicates that a hit song has a strong and sustained growth trajectory, where each listener contributes to a steady and long-lasting increase in popularity.

Figure 7 shows the distribution of the reproduction number for both viral and hit songs. The median values are 3.769 for viral and 3.478 for hit songs, meaning that viral songs spread at a slightly higher rate than hit songs. However, while there is a statistically significant difference between the two distributions, the effect size is relatively small, as the p-value is not particularly low. This suggests that, although viral songs generally spread more efficiently, the distinction between viral and hit songs regarding reproduction potential is not as substantial as other factors such as the infectious period. This result reinforces the idea that while virality is characterized by rapid adoption, long-term success depends on additional dynamics beyond just the initial spread.

5. Comparison and Discussion

Following the results presented in the previous section, we now discuss how epidemic models, particularly the SIR model used in this work, can represent music virality/success. Such a discussion aims to deepen the understanding of the applicability of such models to the distinct facets of musical popularity, ultimately addressing our research question (i.e.,

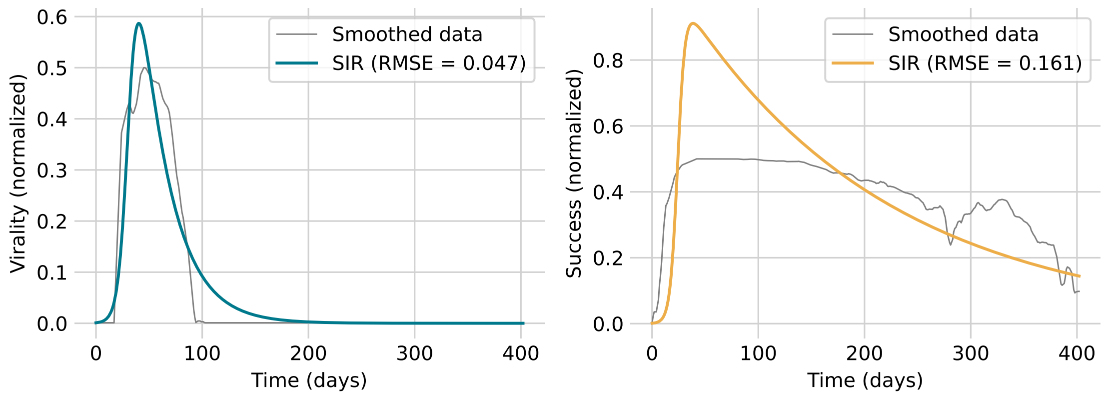

Figure 8. Virality (left) and success (right) time series with their respective SIR model fits for the song "Batom de Cereja - Ao Vivo" by Israel & Rodolffo.

can music popularity in Brazil be effectively modeled as an epidemic process?).

Regarding the overall fitting results, it is reasonable to say that the SIR model provides a better fit for virality curves than for success curves. This may be justified by the fact that viral songs present a rapid increase and decline in popularity, which aligns well with the infection and recovery dynamics modeled by SIR. In contrast, hit songs tend to sustain their popularity over longer periods, a pattern that is not captured by the SIR framework as well. This suggests that while this model may be effective for modeling short-term, fast-spreading trends, it may require adaptations or alternative approaches to represent long-term success accurately.

An example that illustrates such a difference is the song "Batom de Cereja" by the *sertanejo* duo Israel & Rodolffo. Initially, it spread quickly through social media, which was leveraged by the fact that one of them was in the cast of a very popular reality show in Brazil. Such a spread is reflected in its virality time series (Figure 8, left), which resembles a process of viral contagion that the SIR model captures well. However, it also managed to transform its virality on streams, which is depicted by its success time series (Figure 8, right). This longer type of popularity is particularly not well-captured by the SIR framework, as confirmed by the higher RMSE value.

The differences between viral and hit songs are also reflected in the SIR model parameters. The infection rate β is higher for viral songs, reinforcing that they spread more quickly. Meanwhile, the recovery rate γ is also higher for viral songs, indicating that their decline is faster than for hit songs. This result aligns with previous work on music popularity dynamics, reinforcing that virality and success follow distinct diffusion patterns: viral songs experience ephemeral popularity spikes, whereas hit songs maintain a more enduring growth trajectory [Oliveira et al. 2024a].

Failure cases. Despite its strengths, the SIR model faces challenges in capturing songs that experience a resurgence in popularity after an initial decline, such as "Shallow" by Lady Gaga and Bradley Cooper (Figure 9). Since SIR assumes a strictly decreasing recovery phase, it fails to represent cases where a song re-emerges due to external factors such as social media trends or strategic marketing campaigns. Such a finding highlights the need for models that can fit more complex virality/success dynamics, such as periodic

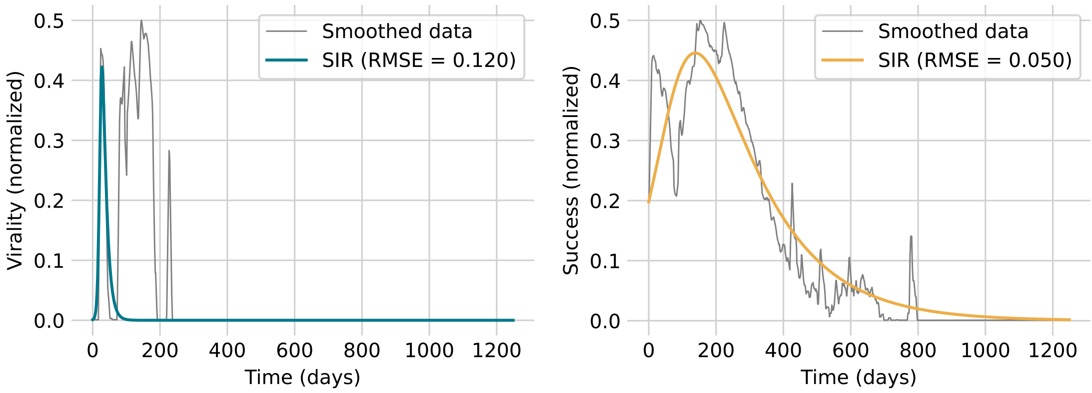

Figure 9. Virality (left) and success (right) time series with their respective SIR model fits for the song "Shallow" by Lady Gaga and Bradley Cooper.

revivals or even multiple periods of popularity.

6. Concluding Remarks

In this work, we investigated whether epidemic models can effectively represent the dynamics of music popularity in Brazil. By applying the Susceptible-Infected-Recovered (SIR) model to both songs' virality and success trajectories, we analyzed how those two different types of popularity evolve over time. Besides evaluating the fitting results, we also examined how the SIR parameters can be interpreted in the context of music popularity. The results show that SIR is useful in capturing how songs become popular, especially for virality. The higher infection rates and shorter infectious periods reflect the fast social media-driven trends for viral songs, whereas hit songs follow a more stable trajectory, making them less similar to a viral epidemic, and therefore less suited to this model.

Overall, such results allow answering our research question because music popularity in Brazil can be modeled as an epidemic process to some extent. The fact that the SIR model fits better for virality when compared to success corroborates the idea that both virality and success are different but connected sides of a song's popularity. Therefore, it is necessary to use different modeling approaches to represent such distinct processes accurately. All such insights contribute to a deeper understanding of how music consumption is shaped in the streaming age, in which the rapid dissemination of content through social networks is increasingly central to the music industry's dynamics.

Limitations and Future Work. The main limitation of this work is the reliance on Spotify data, which may not fully capture the broader dynamics of music popularity across different social platforms. For example, the absence of an exact formula for calculating the songs' virality may lead to an incomplete understanding of how the songs are being shared. Furthermore, while the SIR model offers valuable insights into music virality and success, future work should address more complex models that capture cases in which a song has multiple periods of popularity or re-emerges after a significant decrease in it.

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

References

Barbieri, N. and Bonchi, F. (2014). Influence maximization with viral product design. In *SDM*, volume 1, page 55 – 63.

Barbosa, G. R. G. et al. (2021). Hot Streaks in the Brazilian Music Market: A Comparison Between Physical and Digital Eras. In *SBCM*, pages 152–159. SBC.

Bjørnstad, O. N. et al. (2020). Modeling infectious epidemics. *Nature methods*, 17(5):455–457.

Cançado, P. G. N. et al. (2024). Ajuste de parâmetros do modelo epidemiológico SIR em redes sociais geradas por modelos para simulação da covid-19 em município mineiro. In *BraSNAM*, pages 103–115. SBC.

Centola, D. and Macy, M. (2007). Complex contagions and the weakness of long ties. *American Journal of Sociology*, 113(3):702–734.

Cha, M. et al. (2012). The spread of media content through blogs. *Social Network Analysis and Mining*, 2(3):249 – 264.

Compte, D. L. and Klug, D. (2021). "it's viral!" - A study of the behaviors, practices, and motivations of tiktok users and social activism. In *CSCW Companion*, pages 108–111, Online. ACM.

Dhanaraj, R. and Logan, B. (2005). Automatic prediction of hit songs. In *ISMIR*, pages 488–491, London, UK. ISMIR.

Guerini, M. et al. (2011). Exploring text virality in social networks. In *ICWSM*, pages 506–509, Barcelona, Spain. The AAAI Press.

Li, Y. and Shao, L. (2024). Using an epidemiological model to explore the interplay between sharing and advertising in viral videos. *Scientific Reports*, 14(1).

Ling, C. et al. (2022). Slapping cats, bopping heads, and oreo shakes: Understanding indicators of virality in tiktok short videos. In *WebSci*, pages 164–173. ACM.

Mann, H. B. and Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The annals of math. statistics*, pages 50–60.

Medina, F. J. L. (2023). Social learning and the complex contagion of political symbols in twitter: The case of the yellow ribbon in catalonia. *Big Data Soc.*, 10(2).

Mondelli, M. L. B. et al. (2018). O que os países escutam: Analisando a rede de gêneros musicais ao redor do mundo. In *BraSNAM*, Natal.

Mønsted, B. et al. (2017). Evidence of complex contagion of information in social media: An experiment using twitter bots. *PloS one*, 12(9):e0184148.

Moura, F. A. S. et al. (2024). Characterization of the Brazilian musical landscape: a study of regional preferences based on the Spotify charts. In *WebMedia*, pages 80–88. SBC.

Nika, M. et al. (2015). Going multi-viral: Synthedemic modelling of internet-based spreading phenomena. In *VALUETOOLS*, page 50 – 57.

Oliveira, G. P. et al. (2024a). A quantitative comparison of viral and hit songs in the Brazilian music market. *Vórtex Music Journal*, 12:1–29.

Oliveira, G. P. et al. (2024b). Analyzing the temporal relation between virality and success in the Brazilian music market. In *SEMISH*, pages 157–168. SBC.

Pereira, F. S. F. et al. (2018). That's my jam! uma análise temporal sobre a evolução das preferências dos usuários em uma rede social de músicas. In *BraSNAM*, Natal. SBC.

Raponi, S. et al. (2022). Fake news propagation: A review of epidemic models, datasets, and insights. *ACM Trans. Web*, 16(3):12:1–12:34.

Rosati, D. P. et al. (2021). Modelling song popularity as a contagious process. *Proceedings of the Royal Society A*, 477(2253):20210457.

Sachak-Patwa, R. et al. (2018). Understanding viral video dynamics through an epidemic modelling approach. *Physica A: Statistical Mechanics and its Applications*, 502:416 – 435.

Seufitelli, D. B. et al. (2023a). Hit song science: a comprehensive survey and research directions. *J. New Music Res.*, 52(1):41–72.

Seufitelli, D. B. et al. (2023b). MGD+: An Enhanced Music Genre Dataset with Success-based Networks. In *Dataset Showcase Workshop*, pages 36–47. SBC.

Sharma, R. S. et al. (2011). The emergence of electronic word-of-mouth as a marketing channel for the digital marketplace. *Journal of Information, Information Technology, and Organizations*, 6:41 – 61.

Ternovski, J. and Yasseri, T. (2020). Social complex contagion in music listenership: A natural experiment with 1.3 million participants. *Soc. Networks*, 61:144–152.

Virtanen, P. et al. (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272.