Analyzing the temporal relation between virality and success in the Brazilian music market

Gabriel P. Oliveira, Ana Paula Couto da Silva, Mirella M. Moro

Universidade Federal de Minas Gerais (UFMG) – Belo Horizonte, MG – Brasil

{gabrielpoliveira,ana.coutosilva,mirella}@dcc.ufmg.br

Abstract. *Content virality on social media platforms is essential to modern digital culture. In music, viral songs often gain widespread attention through catchy melodies, relatable lyrics, and captivating visuals. Indeed, social platforms have reshaped music consumption, with viral trends often leading to mainstream success. This study investigates the relationship between music virality and success in Brazil by analyzing their evolution in streaming platforms over time. Through correlation and Granger Causality analyses, we explore the dynamics between these facets of music popularity. Our results show that virality can be used to forecast future success and vice versa, but this cannot be generalized to all songs. Such findings reinforce the differences between the concepts of virality and success besides their symbiotic relationship driven by social platforms.*

1. Introduction

The phenomenon of content virality on the Web, particularly within social media platforms, has become an important aspect of modern digital culture. Such a phenomenon refers to the rapid spread and dissemination of content across various platforms and social networks [Guerini et al. 2011]. For music, a viral song gains extensive attention and popularity due to factors such as catchy melodies and engaging visuals. While music has long been subject to viral spread, platforms such as Twitter, YouTube, and TikTok have significantly magnified this trend by providing reach and accessibility to millions of users worldwide [Kong et al. 2018, Ling et al. 2022].

In recent years, streaming platforms have emerged as key players in music consumption, changing the dynamics of content virality. Platforms such as Spotify, Deezer, and YouTube Music provide access to music produced worldwide, from independent artists to popstars. In addition, platforms such as TikTok have reshaped the way music is discovered, shared, and consumed, with viral trends on the platform often translating into mainstream success for artists. For example, in December 2023, the version of the song "Escrito Nas Estrelas" by Lauana Prado (originally performed by Tetê Espíndola in 1985) topped the charts of the most streamed songs in Brazil after going viral on social media following a contestant's performance on a reality show.[1]

Whereas music success is traditionally related to record sales and radio airplay, the digital era has introduced new metrics centered on streaming numbers [Barbosa et al. 2021]. However, success in the streaming age is distinct from virality, which often manifests as a sudden surge in popularity driven by social media trends. The convergence of streaming and social media has created a symbiotic relationship, where

[1]https://bit.ly/3TFvgVY

viral content fuels streaming numbers, and streaming platforms amplify content virality. In other words, virality can significantly boost the success of a song by launching it to the upper ranks of streaming charts and expanding its exposure to a broader audience.

Within such a dynamic, relevant context, the goal of this work is **to analyze the temporal relation between music virality and success in Brazil**. By studying the trajectories of viral and hit songs in streaming platforms (more specifically, Spotify) over time, we aim to understand the dynamics underlying the interplay between these two facets of music popularity. Specifically, we aim to answer two research questions (RQs):

RQ1. Is there a synchrony between the virality of a song and its success?
RQ2. Can the virality of a song be used as an indicator of its future success or vice-versa?

After reviewing related work (Section 2) and defining virality and success (Section 3), our contributions include: (*i*) building time series for song trajectories on viral and success charts (Section 4); (*ii*) performing correlation analysis to reveal synchrony patterns between virality and success (Section 5); and (*iii*) applying Granger Causality to test if a song's viral trajectory indicates its future success and vice versa (Section 6).

2. Related Work

Investigating the factors that aid musical success is the goal of an emerging field in computer science known as Hit Song Science (HSS). In short, studies within such a field analyze musical data from diverse sources looking for any relationship between a song's features and its success. From the seminal work of [Dhanaraj and Logan 2005], several features have been explored to assess such a goal, including acoustic fingerprints, lyrics, and artist metadata [Seufitelli et al. 2023a].

Moreover, features extracted from artists are important to understanding success. For example, Barbosa et al. [2021] and Oliveira et al. [2023] evaluate the concept of hot streaks (i.e., periods in which the success is above the normal) in musical careers in Brazil and the United States. Collaboration between artists has also been studied as one of the factors that influence musical success [Bryan and Wang 2011, Pereira et al. 2018]. Specifically, Silva and Moro [2019] use Granger Causality analysis to evaluate the temporal statistical relationship between collaboration and success.

Besides being interconnected, success and virality are not the same thing, but rather two distinct faces of music popularity [Oliveira et al. 2024]. Existing research on music virality primarily focuses on viral marketing and listeners' behavior [Kahl and Albers 2013, Fink et al. 2021]. Moreover, Araujo et al. [2019] analyze how a song's presence on viral charts influences its subsequent success. They model such a problem as a classification task, leveraging acoustic features and their viral status to predict its future success and vice versa.

Although our work is closely related to HSS, we do not aim to predict musical success (or virality). In contrast, we aim to investigate the temporal relationship between virality and success using statistical approaches over time series data. In other words, we are not only interested in the status of a song, whether it is a viral or a hit, but we focus on the songs' behaviors as hits and virals over time and how such behaviors are interconnected. Therefore, our goal is to advance the understanding of the music consumption dynamics, especially in a diverse and particular market such as Brazil.

3. Virality and Success Definition

Virality and success represent two distinct yet interconnected facets of music popularity, each uniquely stepping into how songs and artists are recognized and how they impact the industry. Studying such concepts helps to understand the dynamics of music consumption in the digital age [Oliveira et al. 2024]. Here, we summarize the definitions of musical success and virality considered in order to understand the dataset and how we use it.

The concept of **virality** is related to quickly spreading and disseminating content across various platforms and social networks [Guerini et al. 2011]. For music, a viral song gains widespread attention when it is shared by thousands or even millions of users in a very short time span. In contrast, musical **success** is associated with other metrics, including chart performance, streaming numbers, or album sales [Seufitelli et al. 2023a]. Thus, while virality may serve as an entry point to the music industry, success is measured by factors extending beyond mere fleeting trends.

To analyze the temporal dynamics of music popularity in current times (driven by digital consumption), we consider the top-chart perspective for measuring both virality and success. Nowadays, all major streaming platforms (e.g., Spotify, YouTube, and Deezer) produce rankings of viral and hit songs. Therefore, we consider viral all songs that have entered a viral chart, whereas hits are those songs that have made it into a distinct success chart (e.g., the most listened-to songs). We do so regardless of their position in the charts, that is, songs ranked first and last are equally considered viral/hit.

4. Data Preparation

Based on the definitions of viral and hit songs from the previous section, we now present the methodology for both preparing data and analyzing songs. First, we describe an existing dataset and the reasons to use it in Section 4.1. Then, we define the time series modeling that represents the temporal evolution of songs' success and virality on music charts in Section 4.2.

4.1. Dataset

In this work, we consider data from Spotify, one of the world's most popular streaming services, with over 602 million users in 180 markets (as of February 2024).[2] The platform produces the Top 200 and Viral 50 charts, which classify **hit** and **viral** songs, respectively. The former ranks the most listened-to (i.e., streamed) songs, whereas the latter contains the songs gaining the most attention on the platform by considering the rise in plays, sharing, and people who have recently discovered such songs.[3]

In March 2022, there have been significant changes in the Spotify Charts platform, and freely downloading the charts is no longer possible. Therefore, we use the Music Genre Dataset (MGD+) [Seufitelli et al. 2023b], which is based on enhanced data from Spotify Charts. MGD+ contains both global and regional daily charts spanning from January 2017 to March 2022. Here, we consider only data from Brazil, the 9th largest music market in the world in 2022 according to the International Federation of the Phonographic Industry (IFPI).[4] The use of Spotify data for the country is also representative since it is

[2] About Spotify: `https://newsroom.spotify.com/company-info/`

[3] Spotify: `https://bit.ly/3QfZ35W`

[4] IFPI Global Music Report: `https://www.ifpi.org/resources/`

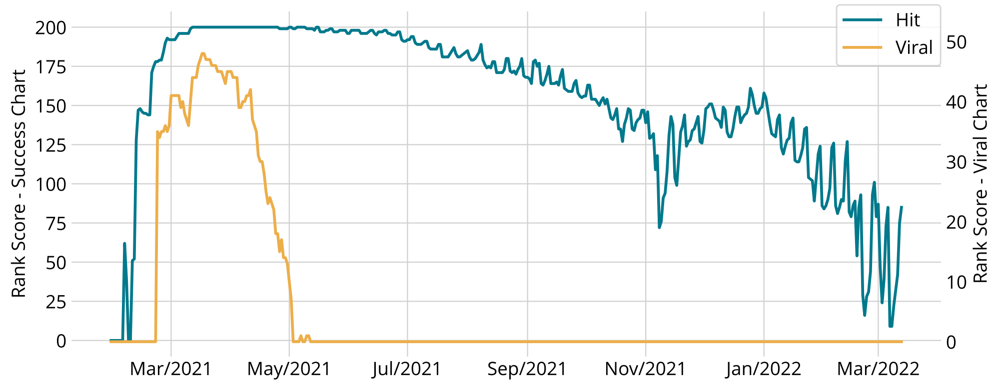


Figure 1. Time series for the song "Batom de Cereja - Ao Vivo" by Israel & Rodolffo. Note the different y-axis scales for success and viral charts.

the most used music streaming app there.[5] Overall, we evaluate 1,895 daily success and viral charts comprising 9,728 distinct songs from 5,126 artists.

Since our goal is to analyze the temporal dynamics of hit and viral songs, with a particular focus on studying whether virality serves as an indicator of success, we refine the dataset to include solely those songs that appear in both charts. In other words, we consider only songs that have achieved hit and viral status. Therefore, our final dataset has 1,977 songs, representing 39.4% and 29.5% of hit and viral songs, respectively.

4.2. Time Series Modeling

Time series analysis has been extensively used in several domains, from metal production [Ramos et al. 2020] to the oil industry [Rossi et al. 2023]. In this work, we create two distinct time series from Spotify charts for all songs to evaluate the temporal evolution of their success and virality, respectively. Such time series serve as input for answering our research questions. Specifically, for each song, we build a time series spanning from its release date (or the first collected chart, if the song was released before 2017) to the most recent chart available. Thus, each data point in the time series represents the song's daily success/virality as indicated by its chart performance.

We then use the rank score to measure such performance, which is calculated based exclusively on the position on the charts [Oliveira et al. 2023]. In short, the *rank_score* of a song of rank i is $rank\_score(i) = max\_rank - i + 1$, where *max_rank* is the lowest possible rank (200 for the success chart and 50 for the viral chart), and i is the position of the song on the chart. For example, if a song reaches the #1 position of the Top 200 charts (i.e., the success chart), its $rank\_score$ is equal to 200 since $rank\_score(1) = 200 - 1 + 1 = 200$. If the song does not reach the charts on a specific day, we set its rank score to zero.

To illustrate our methodology, Figure 1 displays the time series for the song "Batom de Cereja - Ao Vivo" by Israel & Rodolffo. This song was released on February 5, 2021, and was the most streamed on Spotify Brazil that year.[6] On the day of its release,

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

[6]G1: `http://glo.bo/3Tey2Tc`

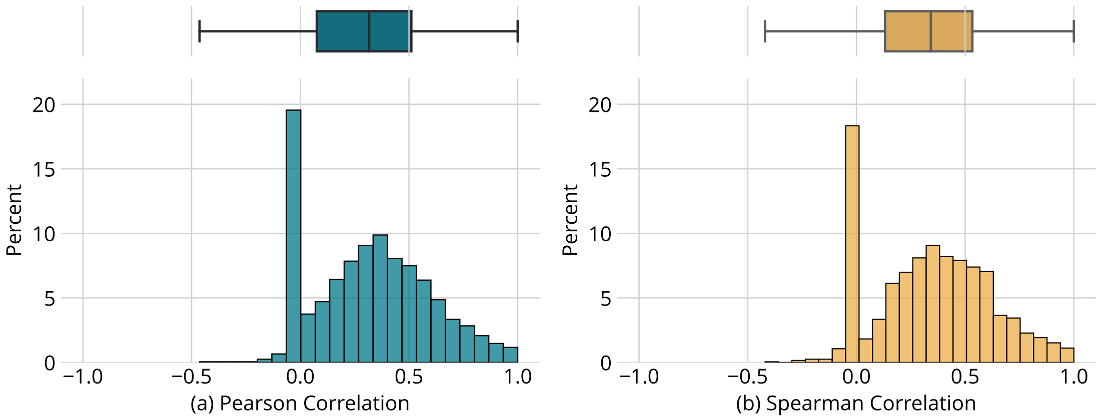


Figure 2. Distribution of the (a) Pearson and (b) Spearman correlation coefficients for the songs in our dataset.

the song debuted at position #139 (i.e., $rank\_score = 62$) on the Top 200 Brazil chart, which ranks the most listened to songs in the country. However, the song quickly climbed several positions, remaining at the top of the charts (i.e., $rank\_score = 200$) for several consecutive days. As for its performance on the Viral 50 Brazil chart, the song entered the chart only on February 22 at position #16 ($rank\_score = 35$). Compared to the success chart, it stayed on the viral chart for a much shorter period of time, reaching its peak at position #3 ($rank\_score = 48$) on March 17.

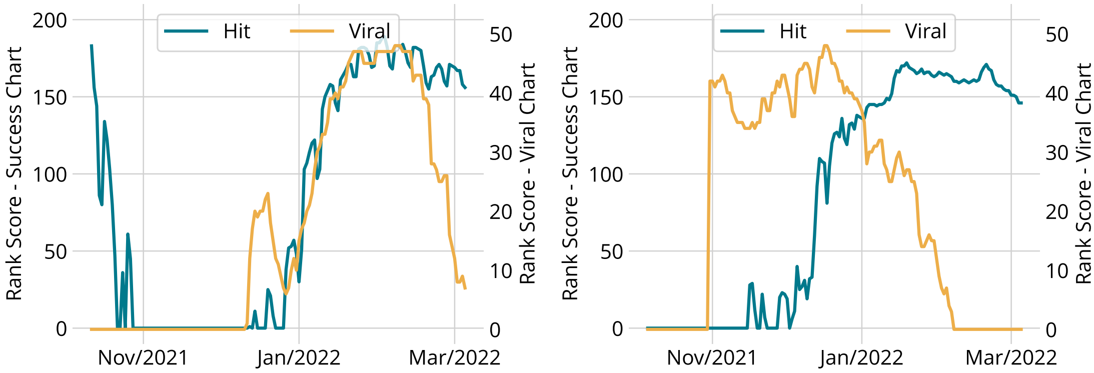
5. Correlation Analysis

To address our first research question (**RQ1**), we analyze the synchrony between the trajectories of song virality and success within the dataset. By analyzing the time series derived from both hit and viral charts, we aim to verify whether there is any correlation between them in order to better understand the relationship between virality and success in the Brazilian music industry. Although simple, correlation analysis is a powerful tool to unveil significant patterns in the temporal evolution of hit and viral songs, offering valuable insights into their dynamic relationships over time.

Here, we calculate two distinct correlation coefficients: Pearson (r) and Spearman (ρ). Whereas Pearson correlation assesses the linear relationship between two variables, Spearman correlation evaluates the monotonic (i.e., rank-based) association between them. Both are numbers ranging from -1 (negatively correlated) to 0 (not correlated) to 1 (perfectly correlated). Thus, we calculate the two coefficients for each song in the dataset by comparing their success and virality time series.

Figure 2 illustrates the distribution of Pearson and Spearman correlation values for the time series. In both cases, a considerable proportion of songs exhibit coefficients close to zero (approximately 20% for Pearson and 18% for Spearman), suggesting that there is no linear or monotonic correlation between success and virality for these songs. Such an observation highlights the complexity and variability in the relationship between these variables across the dataset.

However, most of the songs in the dataset present positive correlation coefficients (around 79.3% of the songs), revealing a tendency (weak or strong) of synchrony between



(a) Idiota - Jão   (b) Assalto Perigoso - Melody

Figure 3. Time series for the songs (a) "Idiota" by Jão and (b) "Assalto Perigoso" by Melody. There are distinct y-axis scales for success (left) and viral (right) charts.

the virality and their success. Specifically, the median correlation value is 0.317 for Pearson and 0.342 for Spearman, representing a weak to moderate correlation (according to Cohen [2013]). Note that while Pearson's coefficient assesses the synchrony between the actual values of the virality/success metric (in our case, the rank score), Spearman's coefficient only considers the order of these values, making it robust against outliers and non-linear relationships. Since both correlation coefficients present similar results, we proceed to analyze only the Pearson Coefficient (r) from now on.

Songs with strong positive correlations. Although the median correlation value is weak to moderate, some songs show a strong correlation between their viral and success trajectories. Specifically, six songs present nearly a total positive correlation ($r \approx 1$), indicating perfect synchrony between such trajectories. A detailed manual analysis reveals that all such songs have a single entry (i.e., they are present in only one day) on both charts. Consequently, the shapes of the two curves mirror each other closely, explaining the extremely high correlation values.

However, there are still other songs with a high correlation between their viral and success time series. As an example, we analyze the song "Idiota" by Jão (Figure 3a), which has a correlation coefficient of $r = 0.806$. The song debuted on the success chart (i.e., Top 200 Brazil) on the same day of its release within the album "Pirata" in October 2021, maintaining its presence for the following days. However, it did not enter the viral charts during this first period. One possible hypothesis for such a behavior is that the song was widely listened to within the artist's dedicated fan base. Indeed, the whole album was highly anticipated by Jão's fans, which led to a robust initial number of streams and chart performance within this established audience.

Despite the initial difference between the two curves, the high correlation is explained by the second period in which the song was present on the charts. This period started in December 2021, when the song entered both the viral and success charts. From then on, the evolution of the song followed a similar trajectory on both charts, reaching its peak between January and February 2022. However, the song experienced a decline in positions on the viral chart shortly thereafter, while maintaining a relatively stable performance on the success chart. Although the time series is interrupted in March 2022, this behavior is in line with the intuition of the difference between the concepts of virality and success, with the former being more ephemeral while the latter is more lasting [Krijestorac et al. 2020, Oliveira et al. 2024].

Songs with negative correlations. Unlike most of the songs in the dataset, a subset of 13 songs shows a representative negative correlation ($r < -0.1$) between their viral and success trajectories. In other words, as a song's position on the viral chart increases, its position on the success chart decreases, and vice versa. An illustrative example is the song "Assalto Perigoso" by Melody. The song was also released in October 2022, and it is a remix of "Positions" by Ariana Grande in a *piseiro*[7] version.

The song gained significant attention on social media platforms, resulting in a large wave of sharing and a great discussion about the song. As depicted in Figure 3b, this viral process helped the song to enter Spotify's viral chart first, maintaining high positions in this ranking for a few months. After a few days, the song also started appearing on the success chart and grew there from then on. The negative correlation observed between these two trajectories shows the phenomenon in which the song's position on the viral charts begins to decline as it ascends within the success chart.

Based on the correlation analysis, we can now answer **RQ1** (*Is there a synchrony between the virality of a song and its success?*). In short, it is not possible to affirm that there is *always* synchrony (i.e., a high positive correlation) between virality and success. Although this is true for a set of songs, other songs present a low correlation between such trajectories. In addition, correlation is only a snapshot of overall synchrony. Therefore, it does not inform the directionality between the two curves, i.e., which signal leads and which follows. To do so, we now perform a Granger causality analysis to better investigate the temporal relationship between music virality and success.

6. Granger Causality Analysis

We now assess **RQ2** by performing a Granger Causality (GC) analysis to verify whether virality can be used to indicate future success or vice versa. GC is a statistical test proposed by Granger [1969] in the context of econometrics for verifying the usage of one variable in forecasting another in time series data with a particular lag. In short, it assesses whether the past values of one time series provide useful information for predicting future values of another time series beyond what can already be predicted from past values of the second time series alone.

To find GC between two time series X and Y, we perform a statistical test to assess whether including lagged values of X as predictors improves the forecasting of Y compared to a model that only includes lagged values of Y as predictors. If the inclusion of lagged values of X significantly improves the forecasting of Y, then we say that X *Granger-causes* Y. Specifically, GC tests the null hypothesis H_0 that X does not *Granger-causes* Y. If H_0 is rejected with a p-value below the predefined threshold, we accept the alternative hypothesis H_1 that X *Granger-causes* Y. In other words, the past values of X contain valuable information for predicting the future behavior of Y beyond what can be explained by the past values of Y alone.

[7] *Piseiro* is a musical genre originated in the Northeastern region of Brazil that blends elements of *forró*, *sertanejo*, and *arrocha*. It often features upbeat rhythms and catchy melodies.

Besides having “causality” in its name, GC does not imply causality in the traditional sense and should not be used to make direct causal inferences. Instead, GC helps analyze potential relationships between variables and is useful in predicting trends. Therefore, in this analysis, we are not verifying a causal relationship between music virality and success, which would require other extensive analyses. Rather, we are studying whether the viral trajectory of a song can serve as an indicator of future success (and vice versa). Next, we first present the possible outcome scenarios from GC analysis (Section 6.1). Then, we present stationarity check and lag definitions (Section 6.2). Finally, we present and discuss the results of GC (Section 6.3).

6.1. Outcome Scenarios

GC is a unidirectional relationship, i.e., the fact that X *Granger-causes* Y does not necessarily imply that Y *Granger-causes* X. Therefore, to assess the impact of virality on musical success and vice versa, we explore four distinct scenarios from the results of GC in the song set, which are detailed next.

S1. ***The song's virality can be used to forecast its success (Viral $\rightarrow$ Success).*** In this scenario, only the viral time series *Granger-causes* the success time series. In other words, changes or fluctuations in the viral time series precede and have a predictive influence on changes or fluctuations in the success time series.

S2. ***The song's success can be used to forecast its virality (Success $\rightarrow$ Viral).*** This is the converse scenario of S1, where only the success time series *Granger-causes* the viral time series. This implies a relationship in which success plays a significant role in influencing the future virality of a song.

S3. ***Both virality and success can be used to forecast each other (Viral $\leftrightarrow$ Success).*** This scenario represents a bidirectional relationship, in which the two time series *Granger-cause* each other. This suggests a mutual influence between virality and success, with each factor impacting the other in a dynamic and interconnected manner.

S4. ***There is no statistical relationship between a song's virality and its success.*** In this scenario, there is no GC relationship between both time series, meaning that changes or fluctuations in the viral time series cannot be used to predict or influence changes in the success time series, and vice versa.

6.2. Stationarity Check and Lag Definition

One prerequisite for employing the Granger Causality (GC) test is ensuring the stationarity of the time series, i.e., they should have a constant mean, variance, and no seasonal component. To verify this, we use two statistical tests: the Augmented Dickey-Fuller (ADF) and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) ones. Essentially, ADF evaluates the null hypothesis H_0 that the series is non-stationary due to a unit root [Fuller 2009], whereas KPSS tests the null hypothesis H_0 that the time series is stationary around a deterministic trend [Kwiatkowski et al. 1992]. In both tests, we set the significance threshold at 0.05. Among the initial 1,977 songs analyzed, only 660 (30.9% of the total) have stationary viral and success time series, being the input GC testing.

The next step is to calculate the causality lag for each song. To do so, we use a VAR (Vector Autoregression) model, which is a statistical model used to analyze the relationship between multiple time series variables. In a VAR model, each variable is

Table 1. Summary of Granger Causality test results (Outcome scenario, number of songs, and respective percentage related to the whole song set).

Scenario	n	%
S1. Viral → Success	1	0.3%
S2. Success → Viral	58	18.8%
S3. Viral ↔ Success	118	38.3%
S4. No Granger Causality relationship	131	42.5%
TOTAL	**308**	**100%**

Table 2. Results of the Granger Causality (GC) test (scenario; song name and artists; the lag orders employed in the GC test; p-value of the GC test, where *, **, and * mean statistical significance at the 0%, 1%, and 5% levels, respectively).**

	Song	Artist	Lag	Viral → Success	Success → Viral
S1	SAD GIRLZ LUV MONEY Remix	Amaarae, Kali Uchis, Moliy	33	< 2.2e-16***	0.058
	Eu Quero e Você Quer	Preta Gil	10	0.392	0.001**
	HAT-TRICK	Djonga	13	0.970	< 2.2e-16***
S2	Meu Melhor Lugar	Fernando & Sorocaba, Luan Santana, Jetlag Music	4	0.965	< 2.2e-16***
	On The Ground	ROSÉ	12	0.950	< 2.2e-16***
	Zombie	The Cranberries	6	0.861	< 2.2e-16***
	áudio de desculpas	Manu Gavassi	19	< 2.2e-16***	< 2.2e-16***
	Delicate	Taylor Swift	53	< 2.2e-16***	< 2.2e-16***
S3	Deus Me Proteja	Chico César	56	< 2.2e-16***	< 2.2e-16***
	Get It Together	Drake, Black Coffee, Jorja Smith	4	< 2.2e-16***	< 2.2e-16***
	Hino do Flamengo	Orquestra e Coro Cid (Flabanda)	8	0.040**	< 2.2e-16***
	Além do Dinheiro	Filipe Ret, Dallass, Ariel Donato	7	0.842	0.933
	Anti-Amor - Ao Vivo	Gustavo Mioto, Jorge & Mateus	9	0.957	0.981
S4	Chantaje	Shakira, Maluma	6	0.980	0.929
	Ousado Amor	Isaias Saad	59	0.928	0.984
	SAOKO	ROSALÍA	3	0.613	0.872

modeled as a linear function of its own lagged values and the lagged values of other variables in the system. Here, lag refers to the number of past time periods included in the model's equation for each variable. To define the number of lags p, we initially estimate an unrestricted VAR(p) model and select the model that minimizes the information criteria of Akaike (AIC), Bayes (BIC), and Hannan-Quinn (HQIC) [Schwarz 1978].

6.3. Results and Discussion

After verifying the stationarity of the time series and setting the causality lag for each song, we proceed with the GC test. For this purpose, we used the *statsmodels* Python library [Seabold and Perktold 2010]. In cases in which the viral and success time series are exactly equal (i.e., mostly songs that have a single entry in the charts), it is not possible to calculate the GC. Thus, the set of songs was reduced to 308 instances, corresponding to 46.6% of the songs with stationary series and 15.6% of the total songs in the dataset.

Table 1 summarizes GC results according to the outcome scenarios presented in Section 6.1. The dataset has songs from all four possible scenarios, suggesting there is no single pattern in the relationship between music virality and success in the Brazilian market. Table 2 contains examples of songs for each considered scenario, including the causality lag value and the p-values for the unidirectional GC tests for *Viral → Success*

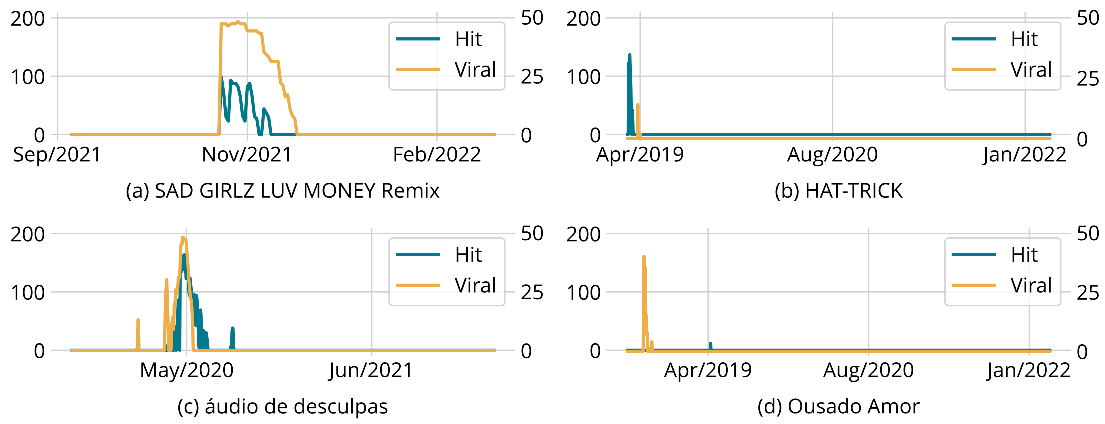


Figure 4. Time series for selected songs in the four scenarios. Note the different y-axis scales for the rank score in success (left) and viral (right) charts.

and *Sucess* $\rightarrow$ *Viral*.

Most songs from the dataset present some form of statistical relationship between success and virality. Regarding scenario S1 (*the song's virality can be used to forecast its success*) only the song "SAD GIRLZ LUV MONEY Remix" by Amararae, Kali Uchis, and Moliy has such a behavior. Its trajectory on both charts presents similar patterns, as illustrated in Figure 4a. Released in September 2021, the remix quickly climbed the viral chart, reaching a peak at #2. Its debut on the success charts occurred 13 days after its entry on the viral charts, reaching a peak position of #102.

The scenario S2 (*the song's success can be used to forecast its virality*) has 58 (18.8%) songs. For instance, the song "HAT-TRICK" by Djonga first reached the success charts and only debuted on the viral charts 13 days later. The song's presence on both charts was equally short, of ten and four days on success and viral charts, respectively. Such a behavior can be justified by an initial surge in streamings driven by the artist's fan base, followed by subsequent virality on social platforms.

Moreover, 118 songs (38.3%) have a bidirectional relationship between virality and success, i.e., scenario S3 (*both virality and success can be used to forecast each other*). An example is the song "áudio de desculpas" by Manu Gavassi, released in 2019, which reached simultaneous peaks of success and virality in early 2020 (Figure 4c). One of the forces driving such peaks of popularity was the singer's participation in a TV reality show, which increased her visibility and made her music reach new audiences.

Among the songs analyzed, 131 (42.5%) have no statistically significant relationship between their viral and success trajectories. This indicates that, for these songs, it cannot be inferred that the viral series *Granger-causes* success, nor that success *Granger-causes* virality (i.e., outcome scenario S4). The song "Ousado Amor" by Isaias Saad is an example of such a scenario, having a large distance between its success and viral curves (Figure 4d). These findings suggest an independence between virality and success, emphasizing the distinct nature of such processes.

Overall, there is not a unique pattern of statistical relationship (using Granger Causality) between virality and success in the Brazilian music market. Recalling our **RQ2**

(*Can the virality of a song be used as an indicator of its future success or vice-versa?*), our findings infer that for some songs, virality can forecast future success and vice versa, but this cannot be generalized to all songs. Thus, our results emphasize the differences between the phenomena of virality and musical success, which despite being two facets of popularity, have significant differences. Furthermore, other factors such as social media discourse and artists' career trajectories before release, which were not considered in this analysis, may directly influence such phenomena.

7. Conclusions

In this work, we analyzed the temporal relation between music virality and success in the Brazilian music market. From a set of popular songs in a digital streaming platform, we built time series for their trajectories in both viral and success charts. We then performed correlation and Granger Causality (GC) analyses to answer our two research questions. The results show it is not possible to affirm there is always synchrony between virality and success. Moreover, the GC analysis corroborated such findings by revealing virality can forecast future success and vice versa for some songs, but for the whole set.

Overall, this study helps to understand the intricate relationship between music virality and success in the Brazilian context. Our findings highlight the multifaceted nature of virality and success as two distinct facets of music popularity, showing that while they often coincide, they do not always correlate uniformly across all songs. Moreover, this work sheds light on the symbiotic interplay between virality and success, in which the influence of social media platforms amplifies the reach and impact of music, reshaping the landscape of music consumption in the digital age.

Limitations and Future Work. The main limitation of this work is that the dataset comprises data from a single digital streaming platform, potentially excluding trends and dynamics from other platforms or offline music consumption. In future work, we plan to include other data sources (i.e., other social platforms) to enhance our analyses.

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

References

Araujo, C. V. S. et al. (2019). Predicting music popularity on streaming platforms. In *SBCM*, pages 141–148, São João del-Rei, Brazil. SBC.

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, Recife, Brazil. SBC.

Bryan, N. J. and Wang, G. (2011). Musical influence network analysis and rank of sample-based music. In *ISMIR*, pages 329–334, Miami, USA. ISMIR.

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Academic press.

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

Fink, L. K. et al. (2021). Viral tunes: changes in musical behaviours and interest in coronamusic predict socio-emotional coping during COVID-19 lockdown. *Humanities and Social Sciences Communications*, 8(1).

Fuller, W. A. (2009). *Introduction to statistical time series*. John Wiley & Sons.

Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, pages 424–438.

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

Kahl, C. and Albers, A. (2013). How to unleash the virus - social networks as a host for viral music marketing. In *CBI*, pages 47–54, Vienna, Austria. IEEE Computer Society.

Kong, Q. et al. (2018). Will this video go viral: Explaining and predicting the popularity of youtube videos. In *WWW (Companion Volume)*, pages 175–178, Lyon, France. ACM.

Krijestorac, H. et al. (2020). Cross-platform spillover effects in consumption of viral content: A quasi-experimental analysis using synthetic controls. *Inf. Syst. Res.*, 31(2):449–472.

Kwiatkowski, D. et al. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of econometrics*, 54(1-3):159–178.

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, Barcelona, Spain. ACM.

Oliveira, G. P. et al. (2023). Hot streaks in the music industry: identifying and characterizing above-average success periods in artists' careers. *Scientometrics*, 128(11):6029–6046.

Oliveira, G. P. et al. (2024). What makes a viral song? Unraveling music virality factors. In *WebSci*, pages 181–190, New York, NY, USA. ACM.

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*. SBC.

Ramos, L. R. et al. (2020). Geração semiautomática de valores de referência para identificação de obstruções em lingotamento contínuo. In *SEMISH*, pages 116–127, Cuiabá, Brazil. SBC.

Rossi, D. F. et al. (2023). Identificação de estáticas em poços de petróleo utilizando motifs. In *SEMISH*, pages 308–319, João Pessoa, Brazil. SBC.

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, pages 461–464.

Seabold, S. and Perktold, J. (2010). Statsmodels: Econometric and Statistical Modeling with Python. In *SciPy*, pages 92–96, Austin, USA. scipy.org.

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.

Silva, M. O. and Moro, M. M. (2019). Causality analysis between collaboration profiles and musical success. In *WebMedia*, pages 369–376, Rio de Janeiro, Brazil. ACM.