Epidemic wave dynamics of music virality on online social platforms

Social Network Analysis and Mining https://doi.org/10.1007/s13278-026-01607-2

Article in Press

Epidemic wave dynamics of music virality on online social platforms

Received: 22 December 2025

Accepted: 26 April 2026

Published online: 31 May 2026

Cite this article as: Oliveira G.P., Vassio L., Couto da Silva A.P. *et al.* Epidemic wave dynamics of music virality on online social platforms. *Soc. Netw. Anal. Min.* (2026). https://doi.org/10.1007/s13278-026-01607-2

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

Epidemic Wave Dynamics of Music Virality on Online Social Platforms

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

[1]Department of Computer Science, Universidade Federal de Minas Gerais, Av. Antônio Carlos, Belo Horizonte, 31270-901, Brazil.
[2]Department of Control and Computer Engineering, Politecnico di Torino, Corso Duca degli Abruzzi, Turin, 10129, Italy.

*Corresponding author(s). E-mail(s): gabrielpoliveira@dcc.ufmg.br;
Contributing authors: luca.vassio@polito.it;
ana.coutosilva@dcc.ufmg.br; mirella@dcc.ufmg.br;

Abstract

Social media and streaming platforms have reshaped music consumption, enabling songs to go viral through social contagion processes that mirror the spread of infectious diseases. In this article, we investigate whether epidemiological models can effectively represent, interpret, and forecast music virality on online platforms. We introduce a wave-based approach that captures multiple independent bursts of popularity, which is not possible using classic epidemic models. We evaluate our approach using data from Spotify and TikTok, comparing its performance with traditional time-series forecasting methods. Our findings demonstrate that epidemic models are significantly more effective at capturing viral dynamics than long-term commercial success. In forecasting tasks, our method achieves accuracy comparable to conventional techniques, even when trained on partial data, while offering interpretable parameters that characterize diffusion speed, engagement duration, and adoption delay. A cross-platform comparison reveals that viral trajectories on TikTok are captured with even greater precision than those on Spotify. Overall, our findings support epidemic modeling as a powerful and interpretable framework for analyzing and forecasting music virality in contemporary digital ecosystems.

Keywords: Epidemiological models, Music virality, Spotify, TikTok

1 Introduction

Social media and streaming platforms have transformed the way music is consumed and shared. Platforms like TikTok and Instagram have become key players in defining which songs rise to stardom, allowing them to reach larger audiences with little or no involvement of traditional media vehicles (Compte and Klug 2021; Ling et al. 2022). However, while such platforms boost song discovery, streaming services are the main channels for music consumption, with recent reports from the International Federation of the Phonographic Industry (IFPI) showing that, in 2023, 73% of people listened to music through streaming services.[1] Furthermore, in 2024, such platforms accounted for 69% of the global recorded music revenue.[2]

The spread of songs across such digital platforms shares many similarities with the propagation of infectious diseases. For example, as diseases spread through contact between individuals, songs spread through social interactions, as users share songs with friends or repost them on social media. In this regard, previous work has applied epidemic models to describe the diffusion of various types of online content. In the music context, such studies investigate the dynamics of song popularity represented as downloads (Nika et al. 2015; Rosati et al. 2021) or views (Li and Shao 2024; Sachak-Patwa et al. 2018).

However, in the digital age, music popularity manifests itself in multiple forms, with virality and success representing two distinct, yet interconnected dimensions of it. Whereas *virality* refers to the fast and often explosive spread of a song through social sharing (Guerini et al. 2011), *success* typically refers to long-term commercial performance, which is usually measured by indicators including (but not limited to) sales and streams (Seufitelli et al. 2023). Therefore, a song can go viral on social media without becoming a widely consumed hit and vice versa.

Given the growing role of social interactions in driving music consumption, modeling the virality and success of music on streaming platforms as contagion processes represents a powerful framework for understanding the complex dynamics of such cultural phenomena. In this work, we investigate whether epidemic models can effectively represent music popularity on social platforms. Specifically, we address the following research questions: **RQ1.** *Are epidemic models suitable for representing music popularity on streaming platforms?* **RQ2.** *How accurately can such models forecast the popularity trajectories of songs?* **RQ3.** *How do the dynamics and performance of epidemic models differ across platforms with distinct virality assessment methods and user engagement behaviors?*

This article extends the paper presented at the 17th International Conference on Advances in Social Network Analysis and Mining (ASONAM 2025) (Oliveira et al. 2025). As a novel contribution, we apply our methodology to TikTok daily video time series to verify the suitability of our proposed approach to other platforms. We use TikTok because it stands out as a key platform where music virality can be observed. In such a platform, music is often consumed passively, as users repeatedly encounter tracks while scrolling through content. This constant exposure enables songs to become

[1] IFPI: https://www.ifpi.org/ifpis-global-study-finds-were-listening-to-more-music-in-more-ways-than-ever/

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

widely recognizable, even without active searching or intentional listening. According to the platform itself, in 2024, 84% of the songs that reached the Billboard Global 200[3] first went viral on TikTok.[4]

Therefore, our main contributions are: (i) we apply epidemic models to songs streaming data from Spotify to capture music popularity dynamics (Section 4); (ii) we introduce a wave-based modeling approach that better reflects the nature of viral diffusion on streaming (Section 5); (iii) we evaluate the forecasting performance of our method against traditional time-series methods (Section 6); (iv) we conduct a cross-platform comparative analysis using TikTok data to investigate the generalizability of our approach and the impact of platform-specific user behaviors on viral diffusion (Section 7); and (v) we perform a qualitative analysis to discuss in depth cases in which our approach fails and succeeds (Section 8). Overall, our results show that epidemic models effectively capture viral dynamics better than success, while achieving forecast performance comparable to conventional approaches. Furthermore, applying our approach to TikTok data also produced promising results, suggesting that the success of this case study represents an important step toward generalizing the modeling of the viral phenomenon as a social contagion process.

2 Related Work

The virality of online content has been extensively studied on different platforms, especially in the context of social networks, where user behaviors and sharing patterns play crucial roles in enhancing its relevance (Compte and Klug 2021; Guerini et al. 2011; Ribeiro 2014) and also in shaping cultural aspects (Duricic et al. 2021). For example, the work of Ling et al. (2022) reveals that virality on TikTok is related to factors beyond the users' followers, including the presence of text and point of view on videos. In addition, Galante et al. (2025) model influencer popularity by combining user activity, content expertise, exogenous events, and platform-driven visibility. Indeed, social media has amplified virality by enabling fast and widespread content sharing. Understanding such dynamics has direct applications in marketing (Castiglione et al. 2021), fighting misinformation (Kumar and Jha 2022), and addressing other social issues (Chou et al. 2020; Mathew et al. 2019).

Specifically in music, interest in understanding the mechanisms behind the increase in popularity of a song resulted in a new research area known as Hit Song Science (HSS). Studies in such a field consider different perspectives to model musical popularity, from chart performance to engagement metrics (Seufitelli et al. 2023). However, music popularity can be understood as a broader concept with distinct facets, namely success and virality (Oliveira et al. 2024a,b). While success is more related to long-term commercial relevance and is the subject of study of HSS, music virality refers to the fast and widespread circulation of a song, being associated with concepts such as word-of-mouth (Sharma et al. 2011) and diffusion processes (Rosati et al. 2021).

[3]The Billboard Global 200 is a weekly chart that ranks the world's best-performing songs based on a combination of digital sales and streaming data from over 200 territories worldwide.
[4]https://newsroom.tiktok.com/tiktok-and-luminate-release-latest-music-impact-report

This fast-paced circulation is significantly amplified by contemporary social media ecosystems, in which TikTok has emerged as a key catalyst for music virality, particularly in the post-pandemic era. Through its short-form and highly shareable video format, the platform enables users to generate and disseminate creative content efficiently, facilitating the emergence of challenges and memes that propel tracks into the mainstream. Indeed, much of this amplification is driven by algorithmic processes and recommendation systems (e.g., TikTok's "for you" page), playing a key role in driving content diffusion beyond users' immediate social connections (Cullen et al. 2025; Baumann et al. 2026). In this context, Coulter (2022) provides an in-depth analysis of how TikTok distinguishes itself from traditional channels in promoting new music releases. Furthermore, Biasioli (2024) investigates phenomena adjacent to viral spread, such as the "memefication" of musical styles and the process through which users reappropriate or remove original meanings in favor of self-expression.

Following the work of Centola and Macy (2007), diffusion online has been addressed as a contagion process, in which individuals adopt a specific behavior after being exposed to it. Such processes resemble the spread of infectious diseases, and epidemic models offer a powerful framework to study their spread over time (Bjørnstad et al. 2020a,b). Although self-exciting point-process models, such as Hawkes processes, are often used to model multiple popularity spikes (Rizoiu et al. 2017, 2018), epidemic models provide a robust and interpretable macroscopic view of the population dynamics. These models have been adapted to digital contexts, from the diffusion of online narratives and information (Gurung et al. 2024; Sivaraman et al. 2023) to the spread of toxicity in online platforms (Addai et al. 2024). In the music context, epidemic approaches have been used to model the temporal dynamics of song popularity by analyzing song downloads (Nika et al. 2015; Rosati et al. 2021) and video views (Li and Shao 2024; Sachak-Patwa et al. 2018).

Unlike previous studies that rely on video views or downloads to assess music virality, to the best of our knowledge, we are the first to leverage temporal streaming data to represent it, which more accurately captures current music consumption patterns. Moreover, we explicitly distinguish between virality and success, considering them as separate but complementary dimensions of music popularity. Regarding the methodology, in addition to employing traditional epidemic models (i.e., SIR, SEIR, and SEIRS), we introduce a novel wave-based approach to identify and characterize multiple independent virality spikes that a song may have over time.

3 Data and Time Series Modeling

We consider Spotify data to represent music popularity, as it is the most used audio streaming service with more than 640 million users over 180 markets.[5] For each market (and for the Global aggregate), the platform produces Top 200 and Viral 50 charts, which are daily distinct song charts that we use to measure music success and virality, respectively. The first is the ranking of the most-streamed songs on the platform, whereas the latter contains the songs gaining the most buzz. Spotify ranks the viral songs based on an undisclosured combination of the increasing rate in plays, the speed

[5]As of January 2025. https://newsroom.spotify.com/company-info/

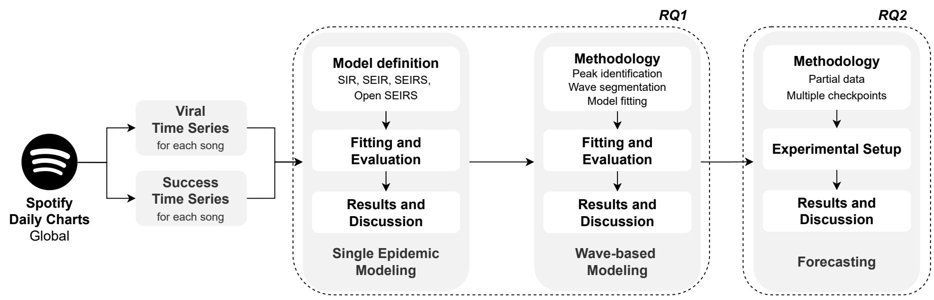


Fig. 1: Overview of the analysis conducted using our data and time series modeling.

of sharing, and the people who discovered them.[6] In this work, we use 1,895 daily viral and success charts for the Global aggregated market, comprising the period from January 2017 to March 2022.[7]

From the daily charts, we build a time series for each song starting from its release date to March 12, 2022, where each point reflects its chart-based popularity on Spotify. Following our previous work (Oliveira et al. 2024b), we use the rank score to measure performance, calculated as $rank\_score(i) = max\_rank - i + 1$, where i is the chart position and max_rank is 200 (success chart) or 50 (viral chart). If a song does not reach the chart, we set the score to zero. We then apply min-max normalization to rescale each series into the $[0, 0.5]$ interval, assuming a maximum of 50% of the population can be infected at once in our models (see the next section). For computational reasons, we set the zero rank scores to a small value (i.e., 0.001), meaning that the song still had some popularity even if it did not reach the charts.[8]

Next, inspired by the work of Sachak-Patwa et al. (2018), we take a simple moving average of seven days for each time series to reduce the noise and smooth the fluctuations in popularity. Moreover, we only consider songs that were present in the charts for more than a week. Hence, our dataset contains 1,647 viral and 1,725 successful songs, with an overlap between the two sets, as some songs appear at least once on both charts. Figure 1 summarizes the analyses performed from such data.

4 Single Epidemic Modeling

In this section, we address RQ1 (*Are epidemic models suitable for representing music popularity on streaming platforms?*) by modeling music popularity using single epidemic models. In other words, we consider the song's viral/success trajectory as a single epidemic process. Following prior work on video popularity (Li and Shao 2024; Rosati et al. 2021), we consider compartmental models to represent such a phenomenon. We aim to verify whether this modeling type can reflect music popularity and, if yes, what is the best model for it.

[6]Spotify: https://support.spotify.com/us/artists/article/charts/
[7]On March 2022, Spotify Charts changed its platform, and it was no longer possible to download the CSV files with the charts.
[8]Other normalization strategies are also possible, such as adding the noise first and then scaling by the maximum possible rank score.

Table 1: Summary of the model states and parameters used in this work.

Term	Description	SIR	SEIR	SEIRS	Open SEIRS
	States				
S	Susceptible individuals	✓	✓	✓	✓
E	Exposed individuals		✓	✓	✓
I	Infected individuals	✓	✓	✓	✓
R	Recovered individuals	✓	✓	✓	✓
	Parameters				
β	Infection rate	✓	✓	✓	✓
γ	Recovery rate	✓	✓	✓	✓
σ	Latency rate		✓	✓	✓
ω	Loss-of-immunity rate			✓	✓
μ	Birth/death rate				✓
α	"Death due to infection" rate				✓

4.1 Model Definition

We consider four different epidemic models for our music popularity time series: SIR, SEIR, SEIRS, and Open SEIRS (Bjørnstad et al. 2020a,b). Although our time series do not directly measure the number of individuals impacted by a song, they can reasonably be considered as proxies for the infection curve in such models. Regarding notation, each model is described by a system of differential equations, in which each equation represents the variation of a state over time t. We use italic uppercase letters to denote the model compartments (i.e., the possible states of individuals in the system) and the total population N. Moreover, the parameters representing the rate of change of each state are denoted by lowercase Greek letters. Table 1 summarizes the states and parameters from all models.

SIR model. This model considers a three-state epidemic, in which individuals can be either susceptible (S), infected (I), or recovered (R). This model considers a fixed population of $N = S + I + R$ and a closed epidemic, i.e., no births or deaths. Here, susceptible means users who have not been exposed to a given song, but may do so in the future. In contrast, infected individuals are those who are actively contributing to the spread of a song by streaming (for success) or sharing (for virality). Finally, the recovered state means that a person who has lost interest in a song and stopped consuming it. The number of individuals in each state is a function of time t, with transitions from susceptible to infected occurring at rate β, and from infected to recovered at rate γ. The change rates of each state are given by Equations 1, 2, and 3.

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

SEIR model. This model builds upon SIR by adding a new state E between susceptible and infected, in which people are exposed before being actually infected. In our context, an exposed individual means someone who has encountered the song indirectly but has not yet actively engaged with it. In other words, this additional state allows capturing the delay between initial exposure and active engagement. Individuals transition from the exposed to the infected state at a rate σ, while the rest of the dynamics follow similar principles to the SIR model. The rate of change for each state is defined by Equations 4, 5, 6, and 7.

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -\frac{\beta SI}{N} \tag{4}$$

$$\frac{\mathrm{d}E}{\mathrm{d}t} = \frac{\beta SI}{N} - \sigma E \tag{5}$$

$$\frac{\mathrm{d}I}{\mathrm{d}t} = \sigma E - \gamma I \tag{6}$$

$$\frac{\mathrm{d}R}{\mathrm{d}t} = \gamma I \tag{7}$$

SEIRS model. It extends SEIR by allowing individuals in the recovered state to return to the susceptible state, introducing the possibility of reinfection. In our context, this reflects the scenario in which users who have previously lost interest in a song may re-engage with it after some time. The transition from recovered to susceptible occurs at a rate ω, and the dynamic transitions between all four states are described by Equations 8, 9, 10, and 11.

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -\frac{\beta SI}{N} + \omega R \tag{8}$$

$$\frac{\mathrm{d}E}{\mathrm{d}t} = \frac{\beta SI}{N} - \sigma E \tag{9}$$

$$\frac{\mathrm{d}I}{\mathrm{d}t} = \sigma E - \gamma I \tag{10}$$

$$\frac{\mathrm{d}R}{\mathrm{d}t} = \gamma I - \omega R \tag{11}$$

Open SEIRS model. It introduces population dynamics into the previous SEIRS model, allowing individuals to enter and exit the system over time. In our context, new listeners join the platform (births), and others become inactive or leave (deaths). The birth and death rates are represented by μ, and while births are only accounted for in the susceptible state, deaths can happen in all of them. There is also an additional "death due to infection" rate α on the infected state, which can represent users who were actively streaming/sharing the song, but permanently disengaged from it due to saturation or shifts in taste. The rest of the dynamics are similar to the simpler SEIRS model (Equations 12, 13, 14, and 15).

$$\frac{\mathrm{d}S}{\mathrm{d}t} = \mu N - \frac{\beta SI}{N} + \omega R - \mu S \tag{12}$$

$$\frac{\mathrm{d}E}{\mathrm{d}t} = \frac{\beta SI}{N} - \sigma E - \mu E \tag{13}$$

$$\frac{\mathrm{d}I}{\mathrm{d}t} = \sigma E - \gamma I - (\mu + \alpha) I \tag{14}$$

$$\frac{\mathrm{d}R}{\mathrm{d}t} = \gamma I - \omega R - \mu R \tag{15}$$

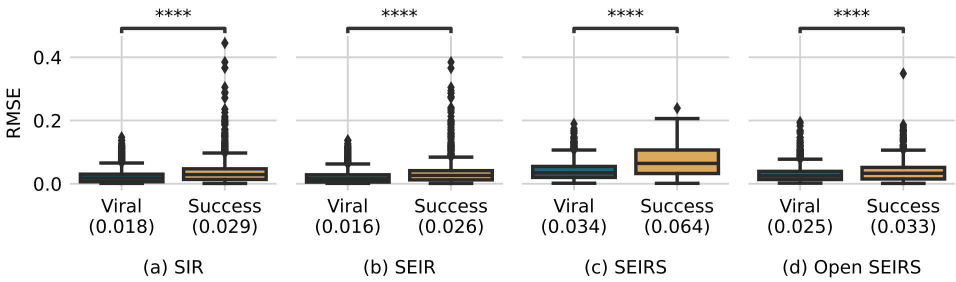


Fig. 2: RMSE for virality and success curves using single epidemic models. Values in parentheses are the median values. Significance is calculated using the Mann-Whitney U test: **** indicates $p \leq 0.0001$.

4.2 Model Fitting and Evaluation

We now define the initial conditions required for fitting the four models to each time series. Since the time series are normalized, we set the total population N to 1.0 in all cases. The initial number of infected individuals, I_0, corresponds to the first observed value in the time series. For all models, the susceptible population at time zero is defined as $S_0 = N - I_0$. All other compartments, i.e., exposed (E_0) and recovered (R_0) are initialized as zero, because we assume that a song's popularity starts with no prior exposure.

We use the *SciPy* Python library[9] to estimate the models' parameters for each popularity time series. We use the least squares approach to perform parameter fitting. For each parameter (i.e., β, γ, σ, ω, μ, α), we set an initial guess of 0.5 and a lower bound of 0 to ensure valid values. Moreover, to evaluate the models' accuracy, we use the Root Mean Squared Error (RMSE) over the whole time period (including when the song is not in the chart), which quantifies the deviation between the observed data and the fitted curve. In our analysis, RMSE values range from 0, indicating a perfect fit, and typically approach 0.5 in poor fits, given the normalized time series. However, since the epidemic model curves are not strictly bounded during fitting, they may exceed the normalized range, meaning there is no fixed upper limit for this metric.

4.3 Results and Discussion

To verify whether epidemic models are suitable for representing music popularity, we evaluate the fitting results for both virality and success time series. Figure 2 illustrates this comparison by showing the distribution of the RMSE values grouped by the four considered epidemic models. The results show that all models performed better for virality time series when compared with the success ones, suggesting that the viral sharing of a song follows a more epidemic-like pattern than its listening behavior measured by streams. In fact, online virality reflects a fast and short-term sharing

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

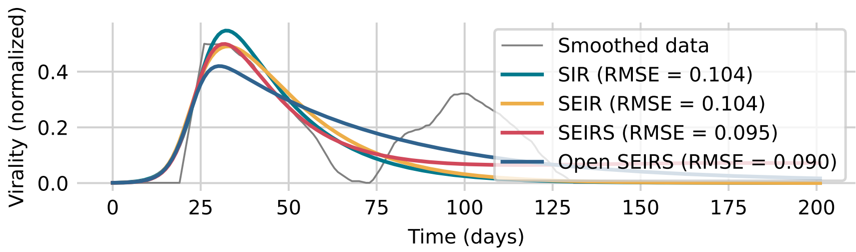


Fig. 3: Model fits for the virality of "Mon Amour - Remix" by Zzoilo and Aitana.

behavior by definition (Guerini et al. 2011), while success may also be related to external factors (e.g., marketing, artist popularity, playlist placement) that the traditional epidemic models do not capture.

Furthermore, when focusing specifically on the virality time series, SEIR produces the best overall fitting performance among the four models considered (median RMSE of 0.016 for virality). Such a finding aligns with the results of previous work on online content popularity, highlighting the SEIR effectiveness in capturing both the initial popularity growth and its subsequent decline (Sachak-Patwa et al. 2018).

However, there are songs for which such models struggle to capture multiple and spaced peaks of virality over time. Even the SEIRS and Open SEIRS models, which allow individuals to be reinfected (i.e., to return to the susceptible state), tend to reach an equilibrium in the long term. An example is "Mon Amour - Remix" by Zzoilo and Aitana (Figure 3), which has two explicit viral moments. However, none of the four models can capture them correctly, highlighting the need for more sophisticated approaches that capture such complex dynamics.

Recalling our RQ1, epidemic models can represent music popularity on streaming platforms to some extent, being more suitable for representing virality rather than success. This reflects a key difference between the two processes: while the first is fast and ephemeral (similar to several infectious diseases), the latter may be longer and influenced by other external factors that are not easily captured by simple epidemic models. However, despite the good results of epidemic models for virality (particularly SEIR), the existence of songs with multiple viral moments opens space for questioning the limitations of the models used and proposing more flexible approaches for representing music virality over time.

5 Wave-based Epidemic Modeling

We now propose a novel wave-based approach to model music virality on streaming platforms, focusing exclusively on virality rather than success, as the latter is not well captured by epidemic models. The central assumption of our approach is that each wave of music virality can be analyzed as a distinct epidemic, potentially leveraged by different factors (e.g., remixes or viral trends). The motivation comes from epidemics such as COVID-19, which unfolded in multiple waves, each driven by a different variant with distinct transmission dynamics.

Note that while we acknowledge the multifaceted nature of music adoption, our approach treats factors such as individual preference as latent variables encapsulated within the model's transmission rates. Similar to biological models in which contact probabilities also aggregate diverse social markers (e.g., distancing or age), we choose to capture the overall population dynamics over granular user categorization.

5.1 Methodology

Our proposed approach is based on the assumption that every song may have multiple virality periods (waves), each one with its own dynamics. Initially, it aims to model and understand the dynamics of music virality, and therefore it works *a posteriori*, i.e., we rely on the complete time series in the fitting process. Our approach is composed of three main steps, described next.

Peak identification. From the preprocessed virality time series, we use a peak detection method[10] to identify significant local maximum points that represent distinct moments of virality. Candidate peaks are local maxima of the time series. To ensure that each peak corresponds to a meaningful and independent event, we define a minimum distance of 30 days between consecutive peaks. If the consecutive candidate peaks are closer than 30 days, we discard the lower ones. The procedure is repeated iteratively over all candidate peaks.

Wave segmentation and adjustment. The peak detection method also returns the left and right base points for each identified virality peak, which we initially consider as the start and end of each virality wave. Such bases correspond to the lowest points surrounding the peak and are determined by scanning outward from the peak until reaching a local minimum on each side. First, we set a minimum width of 7 days for each wave (when $rank\_score > 0.001$), to filter out short-lived spikes that do not represent sustained viral behavior. Moreover, overlaps between waves may occur, especially when peaks are close together. We address this by adopting an independent wave approach where each wave is treated as a separate and self-contained event, which is not affected by adjacent waves. Specifically, when there is an intersection between two waves (i.e., when the right base of the first wave is after the left base of the second one), we shift the starting point of the second wave to immediately follow the end of the first. If the resulting wave has length zero (i.e., it is entirely within the prior), we discard it. We do this because we assume that one wave does not receive any impact from the past, nor does it impact the future.

Epidemic model fitting. Having clearly defined waves, we fit an epidemic model to each one independently. We use the SEIR model based on our analysis of model performance for virality (see Section 4). This is also in line with previous works that state that such a model captures the onset of each wave better (Sachak-Patwa et al. 2018).

[10]We use the `find_peaks` function of the SciPy package: https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.find_peaks.html

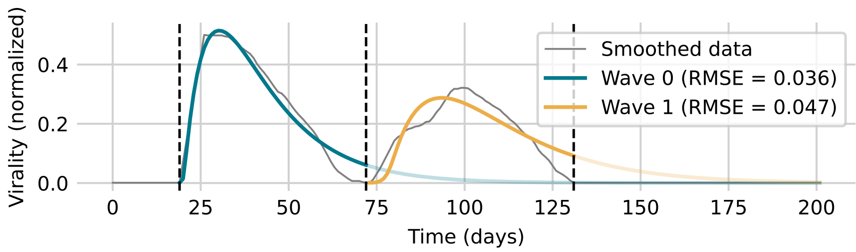


Fig. 4: Virality time series with the wave-based SEIR fit for the song "Mon Amour - Remix" by Zzoilo and Aitana. The vertical dashed lines delimit the waves.

5.2 Fitting and Evaluation

To implement our wave-based approach, we use the `find_peaks` function from the *SciPy* library to identify individual moments of virality within each time series. Once the waves are segmented and adjusted, we fit an independent SEIR model to each wave separately, and the initial conditions follow the same setup described in the single-model approach (Section 4.2).

To evaluate the performance of each wave fit, we compute the RMSE considering only the segment of the time series between the wave's defined start and end points. Such an evaluation allows assessing how well the model captures each individual virality moment. Then, for songs that have multiple waves, we report the overall performance using the average RMSE across all waves.

5.3 Results and Discussion

Since we now fit the SEIR model to each virality wave independently, some songs cannot be considered due to limitations in the model fitting process, namely, the absence of waves after adjustments. As a result, the dataset used for our wave-based analysis comprises 1,045 viral songs (63.4% of the original set). The median value for the average RMSE across all fitted songs is 0.061, indicating a generally good alignment between the model and the observed data. Whereas this value is slightly higher than the median RMSE of the single-model approach, our wave-based method offers a more realistic representation of virality patterns by capturing multiple engagement periods. For example, our approach can capture both virality waves on the song "Mon Amour - Remix", which is not possible using the previous approach (Figure 4).

Regarding the number of waves, the vast majority of songs (994, or approximately 95%) have only one identified wave. Therefore, only a smaller portion has more complex dynamics, with 45 songs having two waves, four songs having three, and only two songs reaching five distinct virality waves. The average and median wave lengths are 38 and 32 days, respectively. Figure 5 presents the distribution of the average RMSE by the number of identified waves. For songs with three and five waves, we show the individual points instead of boxplots due to the small number of samples. In general, there is no statistically significant difference in RMSE across most groups, except for a slight difference between songs with one and two or three waves. However, the median

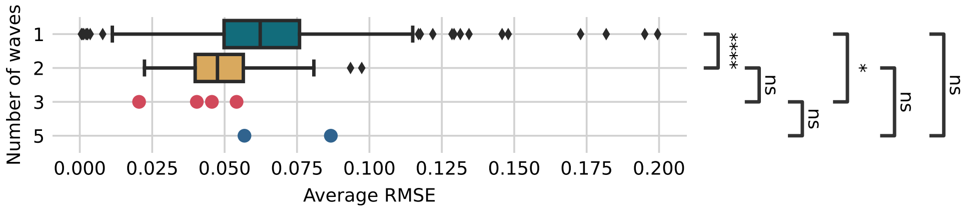


Fig. 5: Average RMSE distribution grouped by the number of identified waves. Significance is calculated using the Mann-Whitney U test: * for $0.01 < p \leq 0.05$; **** for $p \leq 0.0001$; and 'ns' for $p > 0.05$.

Table 2: Descriptive statistics of the SEIR parameters in our wave-based approach.

	SEIR parameters Min.	Mean	Median	Max.
Average infection rate (β)	8.235×10^{-2}	991.970	15.770	1.018×10^{6}
Average recovery rate (γ)	1.605×10^{-7}	0.175	0.113	6.150
Average latency rate (σ)	1.300×10^{-2}	2269.111	0.223	6.271×10^{5}
	Derived parameters Min.	**Mean**	**Median**	**Max.**
Average infectious period ($1/\gamma$)	0.162	1.202×10^{4}	9.181	6.230×10^{6}
Average R_0 (β/γ)	1.092	3.342×10^{4}	161.042	3.353×10^{7}

Table 3: Top 5 songs with highest average infection rate β.

Song	**Artists**	β	γ	σ	**RMSE**
Glorious	Macklemore, Skylar Grey	1.013×10^{6}	0.030	0.013	0.148
Adan y Eva	Paulo Londra	369.256	0.023	0.064	0.087
Dark Red	Steve Lacy	291.631	0.071	0.070	0.090
Notion	The Rare Occasions	240.981	0.031	0.066	0.102
a lot	21 Savage	236.768	0.052	0.068	0.097

values remain relatively close, indicating that our approach maintains a consistent performance even when the number of waves increases.

SEIR Parameters. A major strength of using epidemic models such as SEIR lies in the interpretability of their parameters, which may offer valuable insights into the dynamics of music consumption. Table 2 presents descriptive statistics for the primary SEIR parameters (i.e., β, γ, σ). Since a song may have multiple virality waves, we choose to report the average value of each parameter per song. Such parameters are usually within the range $[0, 1]$, but there is no upper bound since they depend heavily on the shape and scale of each time series.

In our context, the average infection rate (β) measures how fast a song spreads among users, making it a key parameter when analyzing how fast it goes viral. A higher β indicates that a song spreads very quickly in the population, possibly due to strong word-of-mouth combined with marketing strategies. The median value of more than 15 suggests that songs gain traction relatively quickly. However, notice that this may partly reflect: (i) the lack of data before songs enter the Viral 50 chart, limiting our view of early virality growth, and (ii) the normalization to 0.5, likely overestimating the fraction of infected people at the peak. Table 3 contains the five songs with the highest average infection rates. The extreme values happen because all such songs have a high virality rank score at the beginning of the wave, requiring a high β to fit the curve accurately.

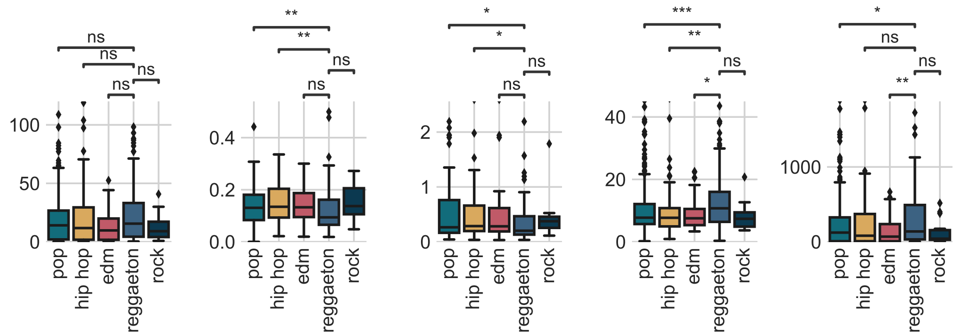
The average recovery rate (γ) reflects how quickly users lose interest in a song once they have started to engage with it, and higher values mean that people lose interest more quickly. The median value of 0.113 suggests that engagement tends to last for a reasonable period before fading. This leads to longer tails, i.e., long periods in the low positions of the chart. In contrast, the average latency rate (σ) captures how fast people move from the exposure to a song to the active engagement with it, with higher values meaning a faster adoption. The median value of 0.223 indicates that, in general, people take some time after being exposed to a song before deciding to share it.

Derived parameters. From the primary SEIR parameters, we can also derive meaningful insights. For example, the infectious period ($1/\gamma$) estimates how long a user stays engaged with a song after discovering it. The median value of around nine days aligns with previous findings that viral songs usually stay popular for a week or two before fading (Oliveira et al. 2024a). Another important parameter is the basic reproduction number ($R_0 = \beta/\gamma$), which represents how many new users a single engaged person is expected to influence. A median R_0 of 161.04 suggests that viral songs have strong contagious potential, reinforcing the idea that music virality is a phenomenon with similar mechanisms to traditional epidemics.

Relation with other music attributes. After characterizing both primary and derived SEIR parameters, we investigate their relationship with other musical attributes. First, we analyze their correlation with acoustic features obtained from Spotify,[11] i.e., descriptors derived from the audio signal itself. Specifically, we consider *acousticness*, *danceability*, *energy*, *instrumentalness*, *liveness*, *loudness*, *speechiness*, *valence*, and *tempo* (Oliveira et al. 2024a). Pearson correlation results indicate that the linear association between SEIR parameters and these features is weak or inexistent, with coefficients ranging between -0.1 and 0.1.

Next, we examine how SEIR parameters vary across selected popular music genres. Since Spotify does not provide genre labels at the track level, we rely on the genres associated with the performing artists. Figure 6 presents the parameter distributions for five genres: pop, hip hop, EDM, reggaeton, and rock. Overall, there are no significant differences across most genres, suggesting that the mechanisms governing music diffusion are largely independent of genre. The main exception is reggaeton, which has statistically significant differences in some cases (e.g., with hip hop and pop for γ

[11] As of March 2026, Spotify no longer provides these features through its Web API.



(a) β (b) γ (c) σ (d) $1/\gamma$ (e) R_0

Fig. 6: SEIR parameter distribution for selected music genres. Significance is given by Mann-Whitney U test: * for $0.01 < p \leq 0.05$; ** for $0.001 < p \leq 0.01$; *** for $0.0001 < p \leq 0.001$; and 'ns' for $p > 0.05$. Unspecified differences are not significant.

and σ). Despite its global popularity, reggaeton retains a strong regional component, which may lead to distinct diffusion dynamics.

Overall, the proposed wave-based approach for modeling music virality complements the findings from the previous section and answers our RQ1 (*Are epidemic models suitable for representing music popularity on streaming platforms?*). Indeed, this approach can effectively represent the dynamics of music virality on streaming platforms, especially when songs have multiple periods of virality. Moreover, given its interpretability and ability to reflect song diffusion patterns, our epidemic approach may also serve as a valuable tool for forecasting music consumption trends, a hypothesis that we explore next.

6 Forecasting Virality Behavior

We now use the wave-based approach to address RQ2 (*How accurately can epidemic models forecast the popularity trajectories of songs?*). Motivated by the parallels between music virality and real-world epidemics such as COVID-19, we explore the potential of our approach to forecast virality trends using only partial time series data. Inspired by prior work on online social dynamics (Ribeiro 2014), we aim to understand whether early virality signals can be used to anticipate a song's future trajectory on streaming platforms.

6.1 Methodology

We design a methodology that operates at the individual wave level to evaluate the SEIR model's forecasting capabilities in the context of music virality. This choice is aligned with our wave-based approach, in which each viral moment is treated independently, mirroring the behavior of separate outbreaks in epidemiological models. For this forecast analysis, we use a dynamic wave-detection heuristic, which analyzes the time series day-by-day without any look-ahead into future data.

In short, a wave onset is considered only after observing five days of consecutive growth in the rank score, while it is declared to have ended when the signal drops below 20% of the local peak reached up to that point. Therefore, we eliminate the data-leakage problem present in the a-priori identification method from the previous section, and we simulate a a scenario in which a song starts gaining traction on social media or streaming platforms, and we want to guess how far and fast it might spread.

To perform this forecasting, we use partial information from the original smoothed virality time series. Specifically, for a given time point t after the beginning of the wave, we use all observed data from the start of the wave up to time t to fit a new SEIR model. This fitted model is then used to predict the subsequent evolution of the wave beyond t. By repeating this process across multiple time checkpoints within each wave, we are able to evaluate how early and how accurately the model can forecast the complete virality pattern.

6.2 Experimental Setup

To evaluate the forecasting performance of the SEIR model, we perform two main experiments at the wave level. In other words, we effectively transform the multi-wave problem into a series of single-wave tasks in which the model is fitted and evaluated. In the first experiment, we simulate different forecast horizons by fitting the model using partial data from the beginning of the wave up to specific cut-off points. Specifically, we consider up to 10%, 25%, 50%, 75%, 90%, and the peak point of the virality curve. With this experiment, we aim to verify the earliest point at which there is a reasonable forecasting and how the performance evolves as more data becomes available.

Following the literature on time series forecasting (Kontopoulou et al. 2023), the second experiment complements the first one and compares SEIR against three baseline forecasting methods: (i) **ARIMA**, a classical statistical model; (ii) **SVR** (Support Vector Regression), which captures non-linear trends; and (iii) **Prophet**, a tool designed for handling time series with seasonality and irregularities (Taylor and Letham 2018). For each baseline, we use the implementation provided by libraries *pmdarima*[12], *scikit-learn*[13], and *prophet*,[14] respectively. Besides having found that SEIR is the model that best fits virality time series (see Section 4), we also consider (v) **SIR** as a baseline, since it is the simplest epidemiological model evaluated in this work. To ensure a consistent and fair comparison of the models' performance, we evaluate the forecasting using RMSE only over the remaining wave portion that was not used during the fitting process (i.e., after the cut-off time t).

6.3 Results and Discussion

The forecasting results are summarized in Figure 7, which presents the RMSE distribution for each fitting data size. As expected, increasing the amount of available data generally improves the forecasting performance, except when considering 90% of the

[12] https://alkaline-ml.com/pmdarima/
[13] https://scikit-learn.org/
[14] https://facebook.github.io/prophet/

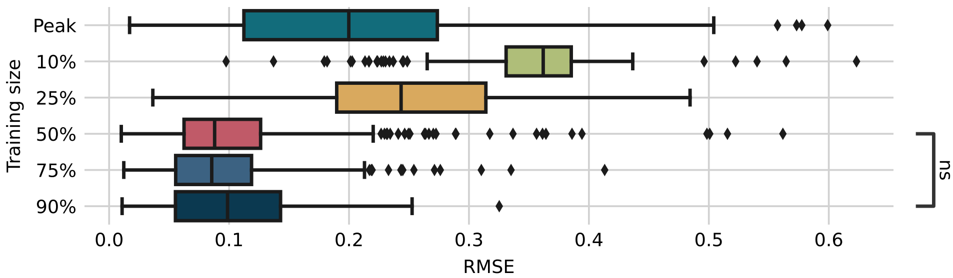


Fig. 7: Forecast RMSE for different partial data sizes. Unless specified with 'ns' ($p > 0.05$), all pairs of distributions are statistically different with $p \leq 0.0001$ (Mann-Whitney U test).

wave. In addition, all pairwise comparisons between training data sizes have statistically significant differences, except the pair between 50% and 90%. Indeed, the peaks in median occur at 31% of the wave time. In addition, using 50% of the wave to fit the SEIR model produces a median RMSE of 0.087, which is only slightly higher than the 0.061 obtained using the whole wave. This suggests that the model can still produce acceptable forecasts of a song's virality even with partial data.

For the comparison with baseline models, we consider only the subset of 391 songs for which all models (SEIR, SIR, ARIMA, SVR, and Prophet) successfully completed the forecast. Figure 8[15] reveals that no single model dominates across all training conditions. However, SEIR performs consistently well, outperforming others at the 50% and 75% marks. In contrast, ARIMA tends to be better when more data is available (especially at 90%), capturing the final gradual virality decays at the end of waves. SVR and Prophet show competitive results in certain conditions, but they generally exhibit higher variability than SEIR. Finally, SEIR outperforms or is statistically equivalent to the simpler SIR model in all scenarios, except when there is more training data available (i.e., 75% and 90%).

Overall, our findings answer RQ2 (*How accurately can such models forecast the popularity trajectories of songs?*), indicating that our epidemic approach can forecast music virality with reasonable accuracy, especially once a wave has begun. Its performance is generally comparable to traditional time-series forecasting methods such as ARIMA, with the advantage of offering interpretable parameters that provide insights into the underlying dynamics of music consumption. For instance, β, γ, and σ reveal how quickly a song spreads, how long users stay engaged, and how fast exposure turns into active engagement.

[15] We set the upper limit of the y-axis to 1 to prevent distortions caused by extreme errors, particularly in Figures 8(a–c), in which Prophet failed to capture the underlying dynamics and produced significantly large forecasting errors.

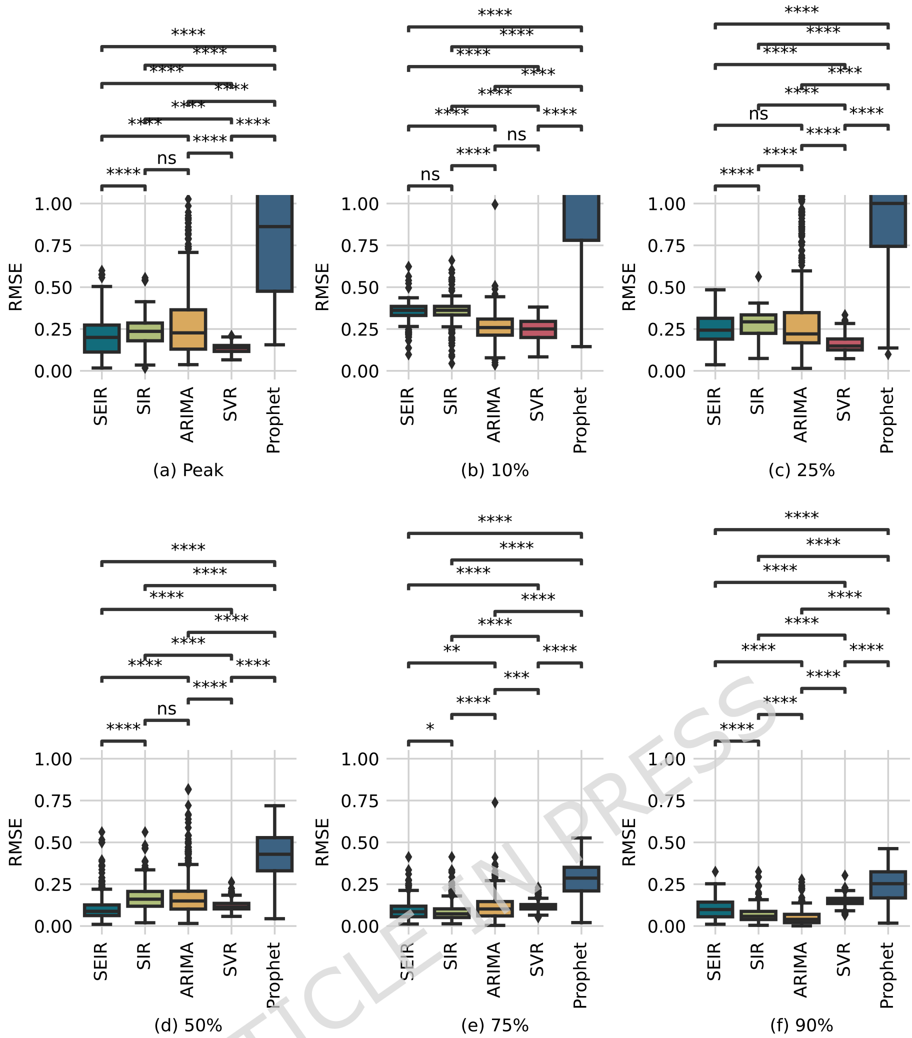


Fig. 8: Forecasting performance for our approach with baselines. Significance is calculated using the Mann-Whitney U test: * for $0.01 < p \leq 0.05$; ** for $0.001 < p \leq 0.01$; *** for $0.0001 < p \leq 0.001$; **** for $p \leq 0.0001$ and 'ns' for $p > 0.05$.

7 Music Virality on Other Platforms: The TikTok case study

In the previous sections, we demonstrated that epidemiological models are suitable for capturing the dynamics of music virality on Spotify. This section investigates whether the same epidemiological modeling can still be used for representing music virality on platforms other than Spotify. Specifically, we assess our RQ3 (*How do the dynamics and performance of epidemic models differ across platforms with distinct virality assessment methods and user engagement behaviors?*) by presenting a case study using TikTok data to replicate our methodology on them and also to assess whether the behavior aligns with our previous findings, thus testing the generalizability of our results.

7.1 Data Collection and Time Series Modeling

To perform this case study on TikTok, we aim to maintain methodological consistency with the analyses presented previously. This alignment allows comparing the performance of epidemiological models when applied to Spotify and TikTok data, ensuring that observed differences are due to platform dynamics rather than methodological variations. Here, we use the TikTok Research API,[16] which provides access to metadata from publicly available videos on the platform. Through this interface, it is possible to retrieve all videos that include a specific song, identified by its corresponding TikTok audio identifier. It is worth noting that, beyond the official audio IDs (typically linked to artists' verified profiles), users may upload other versions of the same song that are later reused in other videos. In this case study, we restrict our analyses to the official audio IDs only. As a result, videos using non-official versions of a song may not be included in our dataset. Furthermore, there may also be songs with few or no videos associated with their official TikTok ID.

Our initial set of songs consists of the 1,045 tracks for which we successfully fitted epidemiological models using Spotify data (see Section 5). However, due to request limits imposed by the API (i.e., 1000 requests and 100,000 records per day), we employ a stratified sampling strategy, selecting 10% of these songs based on their Spotify popularity.[17] This resulted in a subset of 104 tracks. For each of them, we collect the list of short videos published between January 2017 (or the track's release date, if later) and March 2022, in order to match the time span used in the Spotify analyses. In total, our collected dataset contains 4,581,565 videos across all songs, with a median of 4,364 videos per track, ranging from a minimum of six to a maximum of 773,873.

From the collected video lists, we now build time series representing each song's virality on TikTok. Ideally, the most direct measure of this phenomenon would be the number of views over time, as this reflects the number of users exposed to the song. However, the Research API provides only cumulative view counts (i.e., a single static number) rather than the historical evolution of views. Therefore, we represent music virality on TikTok by using the daily number of new videos created with a given track. This proxy captures the dynamics of how often songs are reused by the community and allows for consistency with the temporal modeling framework adopted in this paper.

[16] https://developers.tiktok.com/products/research-api/

[17] The popularity score is retrieved from the Spotify Web API (https://developer.spotify.com/documentation/web-api) and ranges from 0 to 100. Such a variable updates over time and, in our dataset, reflects the collection date (March 2022).

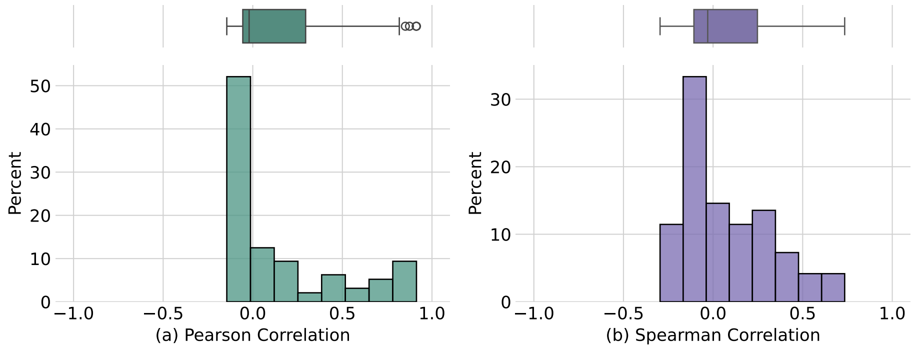


Fig. 9: Distribution of the (a) Pearson and (b) Spearman correlation coefficients for TikTok number of videos and Spotify virality rank score.

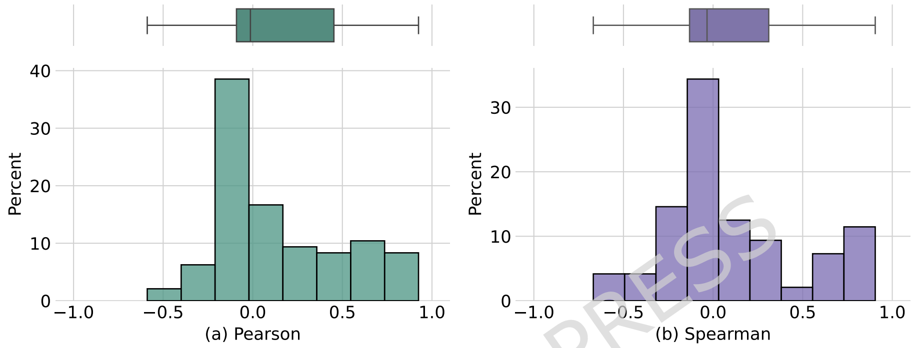


Fig. 10: Distribution of the (a) Pearson and (b) Spearman correlation coefficients for TikTok number of videos and Spotify success rank score.

7.2 Correlation Analysis with Spotify

Before applying our wave-based approach, we first investigate the relation between the number of videos on TikTok and virality on Spotify to assess whether a phenomenon observed on one platform is consistently mirrored on the other. In other words, we test whether a statistical relationship exists between these two dimensions of music popularity. We calculate both Pearson's r and Spearman's ρ coefficients for all songs in our sample, which capture linear and monotonic correlation, respectively.

Figure 9 presents the distribution of correlation values for music virality. Overall, there is no clear linear or monotonic correlation between the TikTok and Spotify time series, with median values of -0.020 (Pearson) and -0.030 (Spearman). For comparison, there is a similar pattern when contrasting TikTok time series with Spotify's success (Figure 10), with medians of -0.013 and -0.033, respectively. Such results suggest that both platforms capture distinct aspects of popularity, potentially reflecting

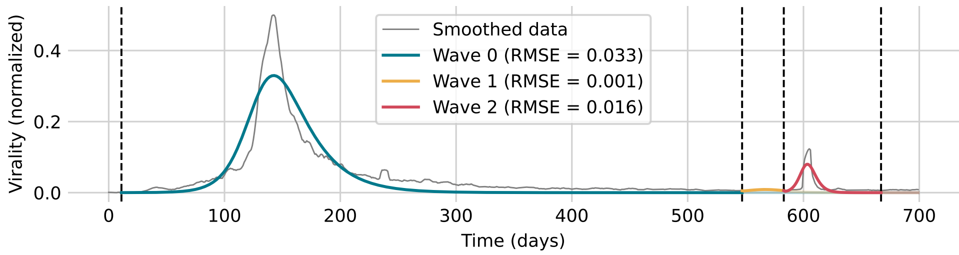


Fig. 11: TikTok virality time series with the wave-based SEIR fit for the song "Heather" by Conan Gray. The vertical dashed lines delimit the waves.

differences in how users interact with music content (e.g., active listening vs. passive exposure or engagement through short videos vs. playlist curation).

Nevertheless, there are songs with strong correlation values between TikTok and Spotify performance. Specifically, 19 songs have significant linear correlations, and seven present significant monotonic correlations (i.e., coefficient greater than 0.5). However, the general lack of significant correlations for most songs may indicate that TikTok and Spotify tend to capture distinct dimensions of the music virality phenomenon. Whereas there are overlaps in certain cases, the broader picture suggests complementary rather than redundant roles. Our findings emphasize the value of examining platform-specific virality dynamics, illustrating how applying epidemic models to TikTok time series can complement analyses conducted on Spotify and other platforms with distinct interaction patterns.

7.3 Wave-based Epidemic Modeling

Now, we apply the wave-based epidemic modeling approach originally proposed for Spotify on Section 5 to TikTok data. By decomposing a song's virality time series into a sequence of independent epidemic waves, we aim to assess whether this methodology can also capture both primary bursts of virality and later revivals on TikTok, thus taking a first step toward the generalization of this approach across platforms.

Our wave-based approach produces a median average RMSE of 0.025, indicating a good alignment between the fitted curves and the observed TikTok data. Also, this methodology allows for a more realistic representation of virality patterns by capturing multiple engagement periods in the platform. For instance, for the song "Heather" by Conan Gray, our approach identified three distinct waves of virality (Figure 11). Although the existence of the second wave may be considered debatable given its relatively small peak compared to the other two, the viral periods are well delimited.

Regarding wave duration, the average and median values are 100 and 45 days, respectively. Such values are larger than those observed for Spotify, which may reflect either platform-specific dynamics or modeling choices in the time series building process. In other words, the TikTok time series provides a continuous measure by

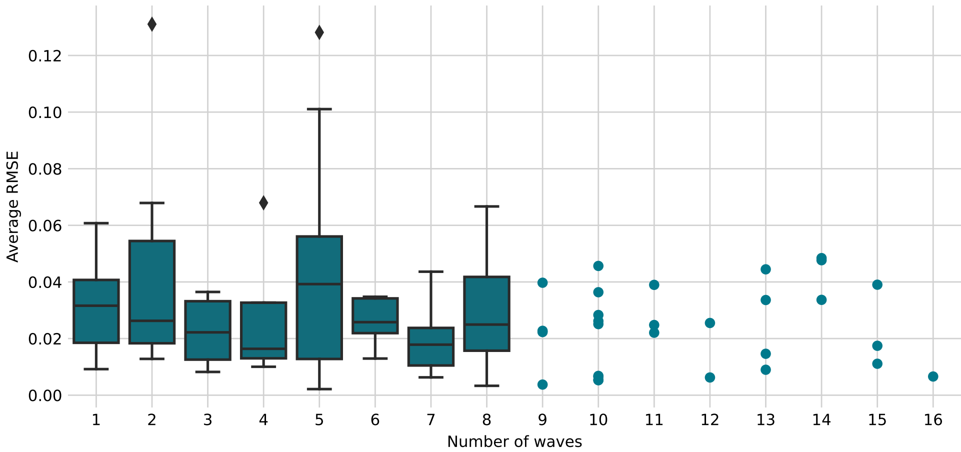


Fig. 12: Average RMSE distribution grouped by the number of identified waves.

Table 4: Descriptive statistics of the SEIR parameters for TikTok time series using our wave-based approach.

	SEIR parameters			
	Min.	**Mean**	**Median**	**Max.**
Average infection rate (β)	0.345	3.617×10^2	4.432	3.294×10^4
Average recovery rate (γ)	8.772×10^{-3}	0.694	0.372	4.185
Average latency rate (σ)	3.062×10^{-3}	1.452×10^5	9.181×10^3	6.233×10^6
	Derived parameters			
	Min.	**Mean**	**Median**	**Max.**
Average infectious period ($1/\gamma$)	0.638	4.036×10^2	12.023	1.804×10^4
Average R_0 (β/γ)	1.109	1.372×10^4	98.039	1.276×10^6

capturing the daily volume of video creations. In contrast, Spotify's visibility is constrained by the charts' inclusion criteria, i.e., data is only accessible when a song reaches the ranking. Hence, songs' virality on days outside the chart remains unknown.

As for the number of waves, TikTok time series have considerably more viral periods than those on Spotify (see Section 5.3). The median is six waves per song, with a minimum of one and a maximum of 16. Figure 12 shows the distribution of average RMSE values grouped by the number of identified waves. For songs with more than nine waves, we show individual points instead of boxplots due to the small sample size. Finally, after applying the Mann-Whitney U test, no statistically significant differences were found between all of the distributions.

SEIR Parameters. We now move to the analysis of each parameter of the SEIR model, which provides meaningful insights into the mechanisms of music virality on

TikTok. Table 4 reports descriptive statistics for the estimated parameters (i.e., β, γ, σ), in which each value is the average across all detected waves of a song. Once again, parameters are generally within the interval $[0, 1]$, but there is no strict upper bound, as the values are determined by the underlying shape and scale of the time series.

The infection rate (β) captures how quickly a song spreads through the platform. On TikTok, the median $\beta = 4.432$ shows that songs can reach new audiences fast, which may be driven by challenges, memes, and content creation. Whereas the distribution includes very large outliers (with maximum values greater than 10^4), these may reflect explosive moments when songs suddenly go viral after being picked up by influential users or trends. However, higher β values could also result from normalizing the time series values to 0.5, which may lead to overestimating the proportion of individuals infected at the peak.

The recovery rate (γ) indicates how fast users lose interest in a song once they are already engaged. With a median of 0.372, this value suggests a considerable time engagement: users do not move instantly on to new content, although they do it faster when compared to Spotify (see Section 5.3). Still, the long tail in the distribution shows that certain songs can maintain attention for longer periods, especially when repurposed in multiple trend contexts.

Moreover, the latency rate (σ) measures how quickly users move from exposure to active engagement. The very high median value ($\sigma = 9.181 \times 10^3$) confirms that adoption on TikTok is almost instantaneous: once a user sees or hears a song, they may reuse it almost immediately in their own content. This aligns with the platform's design, where creative participation (remixing, duets, or challenges) accelerates virality far beyond what is typically seen in other music platforms. It also aligns with findings in other social media contexts, such as Facebook and Instagram, in which the freshness of content is a primary driver of interactions (Vassio et al. 2022).

Derived parameters. The infectious period ($1/\gamma$) has a median of about 12 days, meaning users remain engaged with a song for nearly two weeks before interest fades. In addition, the basic reproduction number ($R_0 = \beta/\gamma$) shows a median of 98, indicating that a single engaged user can, on average, lead to nearly a hundred new adoptions (due to their followers or the exposure of the videos on the "for you" page). Such a high R_0 underscores the explosive viral potential of TikTok, where songs can gain traction in completely new waves of engagement. However, such results should be interpreted with caution, as the normalization applied to the time series may overestimate peak engagement and, consequently, the parameter values.

Cross-platform comparison. Focusing the cross-platform comparison exclusively on the tracks included in the TikTok sample, Figure 13 shows that this method produced lower values of RMSE for TikTok curves, indicating that it better captures the virality dynamics specific to this platform. Furthermore, there are some SEIR parameters with distinct distributions across both platforms, as illustrated in Figure 14. For example, viral curves on Spotify have a higher infection rate (β), suggesting that once a song begins to gain visibility, it may spread more efficiently through recommendation mechanisms and playlist-driven exposure. In contrast, the recovery (γ) and latency (σ) rates are significantly higher on TikTok, which may indicate that user engagement cycles are shorter and that content is consumed and replaced at a faster pace.

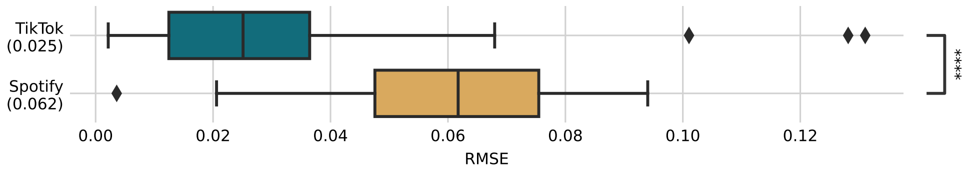


Fig. 13: RMSE for TikTok and Spotify virality curves using the wave-based approach. Values in parentheses are the median values. Significance is calculated using the Mann-Whitney U test: **** for $p \leq 0.0001$.

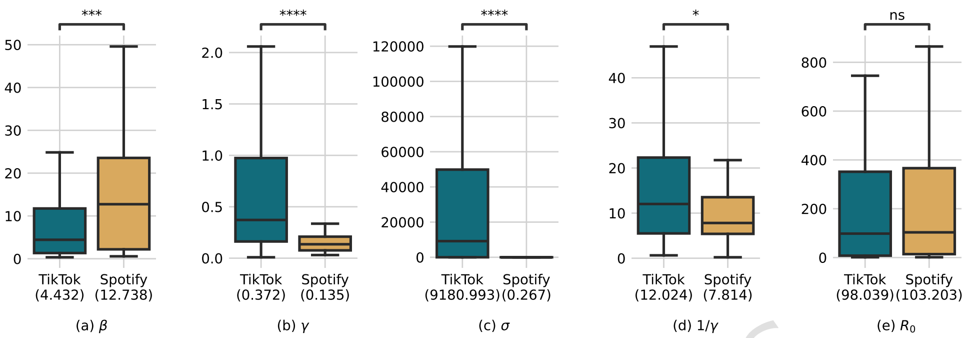


Fig. 14: Parameter values for TikTok and Spotify virality curves using the wave-based approach. Outliers are ommitted for readability purposes. Values in parentheses are the median values. Significance is calculated using the Mann-Whitney U test: * for $0.01 < p \leq 0.05$; *** for $0.0001 < p \leq 0.001$; **** for $p \leq 0.0001$; and 'ns' otherwise.

Regarding the latter parameter, the median σ value is several orders of magnitude higher, suggesting that individuals move almost instantaneously from the exposed to the infected state. In practical terms, this may reflect how quickly users create or interact with new videos once they are exposed to a trend, as well as the potential effects of the temporal granularity and other data-related issues. The infectious period ($1/\gamma$) is slightly longer on TikTok (12 days) than on Spotify (7.8 days). Although this difference is statistically weak, its effect size is significant. Finally, for the basic reproduction number (R_0), no statistical difference was observed between platforms, suggesting that the intrinsic potential for a song to go viral is independent of the platform itself.

7.4 Forecasting Song Virality

Similar to the analysis performed for Spotify on Section 6, we now evaluate our approach's ability to forecast virality trends on TikTok. Specifically, we use the same experimental setup over TikTok time series to investigate whether a song's future trajectory can be anticipated using only its initial performance data on the platform.

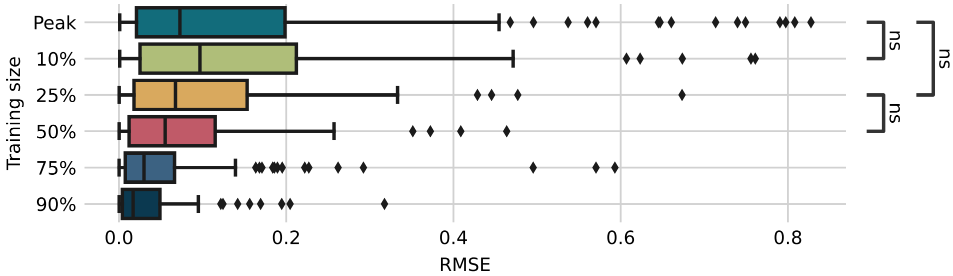


Fig. 15: Forecast RMSE for TikTok with different partial data sizes. Unless specified with 'ns' ($p > 0.05$), all pairs of distributions are statistically different (Mann-Whitney U test).

Figure 15 illustrates the results of the first experiment, which evaluates the impact of the forecast horizon. It shows the evolution of RMSE as different fractions of the time series are used as the model input. Similar to the analysis performed for Spotify, the forecast error decreases as more data becomes available. For example, when using only 75% of the initial wave data, the SEIR model achieves a median RMSE of 0.030, which is very close to the 0.025 obtained when fitting the model with the full viral curve. This finding reinforces that wave-based modeling is robust and effective even with partial data from TikTok, indicating the possibility of predicting a song's viral behavior considerably in advance once a wave has begun.

Moreover, Figure 16 compares the performance of our SEIR-based method with the baseline models for time series forecasting.[18] In general, there is no statistical difference between the models, especially in the 50% and 75% scenarios. In the other scenarios, Prophet is the one with the highest median error. In particular, in the scenario using data up to the wave's peak, it fails significantly, generating RMSE values above 1. In contrast, the error distributions for the remaining models (i.e., SEIR, SIR,[19] ARIMA, and SVR) were quite close, often with no statistically significant differences.

Cross-platform comparison. Figure 17 expands our analyses by directly comparing the SEIR forecasting performance across platforms. Again, we only use the Spotify time series for the songs present in our TikTok sample. The results reveal that our wave-based approach consistently outperforms the models fitted on Spotify data for all forecasting horizons, suggesting that the dynamics captured in TikTok time series are more predictable once the wave structure is partially known. This may be attributed to the distinctive viral mechanisms of TikTok, where engagement patterns follow clearer virality bursts compared to those typically observed in streaming environments.

Overall, the experimental results demonstrate that epidemiological models represent a promising alternative for forecasting the viral behavior of songs on TikTok after

[18]To avoid distortions from extreme errors, we set the y-axis upper limit to 1 in Figures 16(a–c). This adjustment is necessary because Prophet sometimes fails to capture the underlying dynamics, resulting in unusually large forecasting errors.

[19]Similar to Section 6.3, we choose to compare SEIR with SIR even though the latter was proven to produce poorer fit results in Section 4.

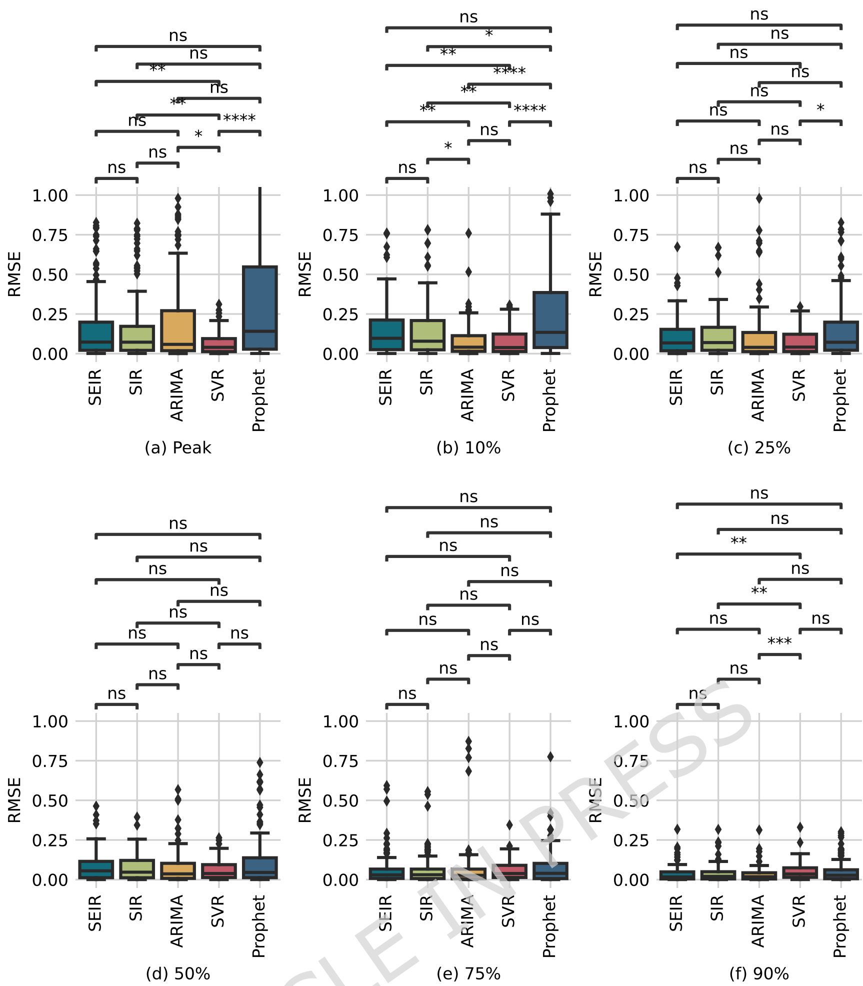


Fig. 16: Forecasting performance for our approach on TikTok data with baselines. Significance is calculated using the Mann-Whitney U test: * for $0.01 < p \leq 0.05$; ** for $0.001 < p \leq 0.01$; *** for $0.0001 < p \leq 0.001$; **** for $p \leq 0.0001$; and 'ns' otherwise.

a wave has started. The SEIR model produces comparable, and in certain scenarios, superior performance to traditional time-series models such as ARIMA. Furthermore, our approach has the advantage of employing interpretable parameters, such as the transmission rate and the infection period, which not only enable forecasting but also reveal intrinsic and important characteristics of the viral phenomenon's dynamics.

Hence, the findings of this section allows answering RQ3 (*How do the dynamics and performance of epidemic models differ across platforms with distinct virality assessment methods and user engagement behaviors?*) by suggesting that epidemic

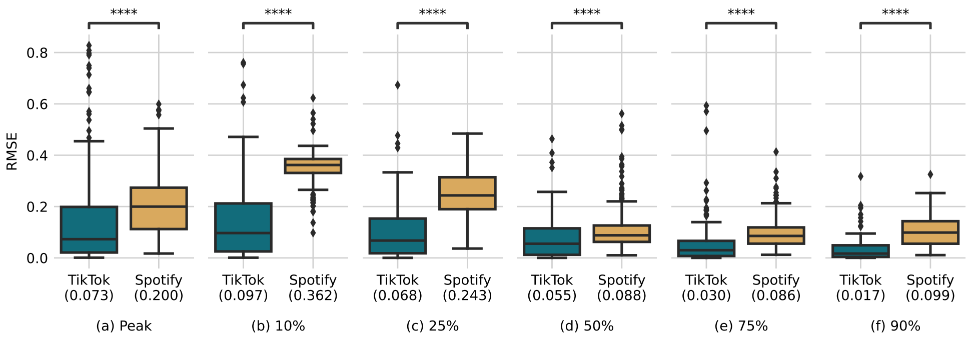


Fig. 17: Forecast RMSE for TikTok and Spotify for different partial data sizes. Significance is calculated using the Mann-Whitney U test: **** for $p \leq 0.0001$.

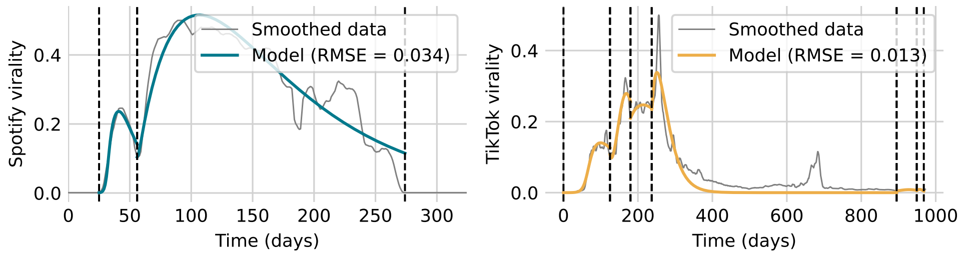
models are versatile enough to represent virality across different digital ecosystems. Our comparative analysis, conducted on the shared subset of tracks, reveals that TikTok viral trajectories are generally captured with greater precision than those on Spotify, although they work well for both platforms.

8 Qualitative Analysis of Music Virality

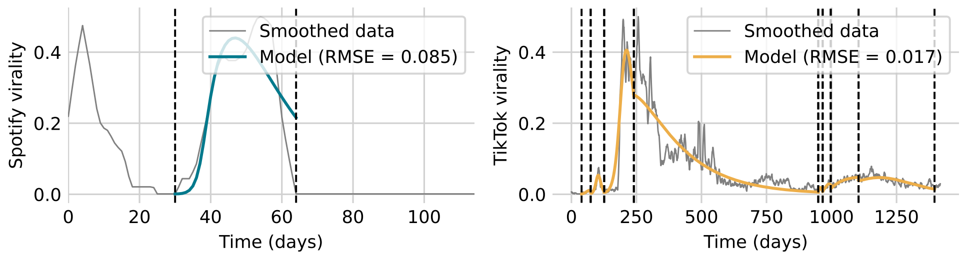
While the previous section quantitatively evaluated the performance of the epidemiological models, this section presents a qualitative analysis of selected songs. The goal is to better understand the diverse patterns of music virality across social platforms and to interpret the factors that may explain the observed variations in model performance. By examining representative songs with distinct model performances, this analysis complements the statistical results and offers interpretative insights into cross-platform music virality dynamics. Therefore, this section is divided into two parts: a cross-platform comparative study (Section 8.1) and an examination of divergent model performances in outlier cases (Section 8.2).

8.1 Cross-Platform Comparative Analysis

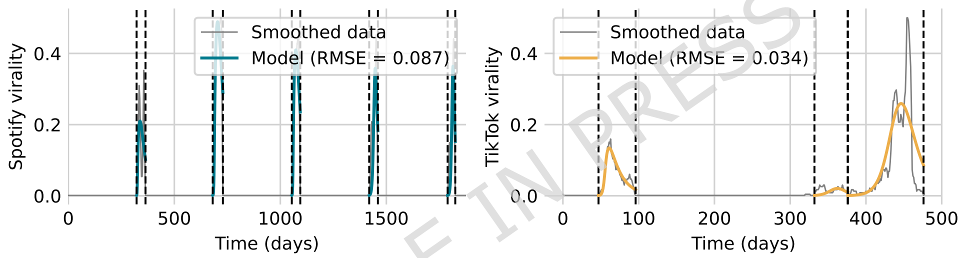
Here, we compare the viral trajectories of the tracks across Spotify and TikTok, highlighting differences in growth patterns, temporal alignment, and model fidelity. To ensure a representative analysis, we employ a sampling strategy based on the average RMSE percentiles of the model fits for both platforms. We classify the songs into three quantile-based error groups according to their average RMSE values: low error (0–25th percentile), medium error (25–75th percentile), and high error (75–100th percentile). From each group, we randomly select one representative song for detailed analysis: "Dance Monkey" by Tones and I (low error), "Youngblood" by 5 Seconds of Summer (medium error), and "Fairytale of New York" by The Pogues featuring Kirsty MacColl (high error).



(a) "Dance Monkey" by Tones and I



(b) "Youngblood" by 5 Seconds of Summer



(c) "Fairytale of New York (feat. Kirsty MacColl)" by The Pogues

Fig. 18: Spotify and TikTok virality time series for selected songs with their respective wave-based model fits. To improve visual clarity, Spotify plots are truncated to 50 days after the end of the last wave.

Analyzing the viral trajectories and their corresponding wave-based model fits, we confirm that the model can capture the fundamental viral movement across different platforms. For instance, for "Dance Monkey" (Figure 18a), which has a low RMSE value, the model accurately captures the main virality waves on both Spotify and TikTok. The largest viral moments happen during the same time window (approximately days 0–300), corresponding to late 2019 and early 2020. In this case, the similarity in the curve shapes suggests a strong coupling between viral exposure on TikTok and sharing on Spotify. Indeed, TikTok has explicitly attributed part of the song's global success to its early virality on the platform,[20] supporting the observed temporal alignment and low modeling error.

For "Youngblood" (Figure 18b), the first peaks of virality on Spotify and TikTok are chronologically close but not simultaneous. On Spotify, the song reaches its main viral peaks shortly after its release in 2018, whereas on TikTok the peak occurs around 200 days later and persists for a considerably longer period. This prolonged viral phase on TikTok may reflect the song's sustained reuse in user-generated content, potentially reflecting the broader commercial success of the track, which Spotify measures through other popularity rankings. The higher fit error observed on Spotify may reflect a specific limitation of our model in capturing the localized peak at the very beginning of the time series.

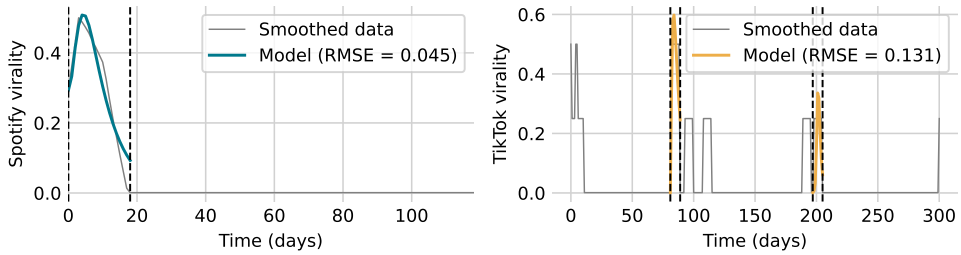
Finally, "Fairytale of New York" illustrates a high-error case with different temporal structures (Figure 18c). A primary discrepancy arises from the time series lengths: although the song was released in 1988, both Spotify and TikTok time series should start on 2017 (see Sections 3 and 7.1). However, TikTok data for this song are only available from 2020 onward, resulting in time series of different lengths and historical coverage. Despite data constraints, the wave-based method still captures the general seasonal structure of the waves, which consistently occur during the Christmas period. Still, the TikTok curves have longer-lasting viral periods, suggesting that seasonal reuse on TikTok extends beyond the consumption peaks observed on Spotify.

Methodological Considerations. It is important to note that RMSE values are directly influenced by the normalization process applied to the time series. Thus, our RMSE values usually range from 0, indicating a perfect fit, and typically approach 0.5 in poor fits. Also, since both Spotify and TikTok data are rescaled to a fixed range, relative differences in curve amplitude and persistence may disproportionately affect error magnitude. Furthermore, the sensitivity of the wave-based approach depends on the parameters adopted during the peak detection and wave segmentation steps described in Section 5.1, including the minimum peak distance and minimum wave width. Such parameters directly affect the identification of viral peaks and the definition of wave boundaries, potentially leading to more conservative or more granular wave decompositions.

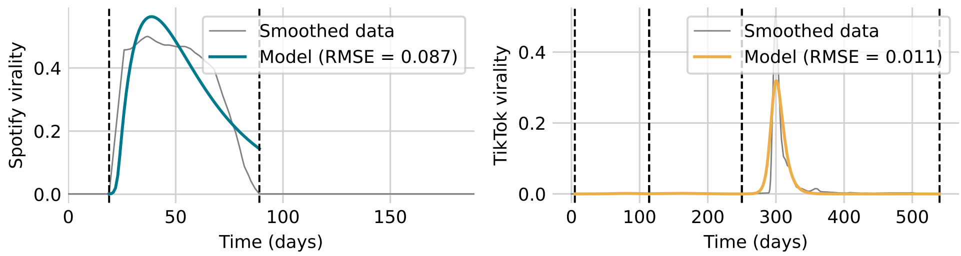
8.2 Divergent Model Performance and Outlier Cases

Now, we identify and discuss outlier songs for which our proposed wave-based approach produces different performance across platforms, i.e., cases in which the model provides a good fit on Spotify but performs poorly on TikTok, and vice versa. Such cases are particularly informative, as they highlight platform-specific limitations that are not fully captured by purely temporal diffusion models. To do so, we select two representative songs: "X (with 2 Chainz & Saudi)" by ScHoolboy Q and "All I Want" by Olivia Rodrigo. Both songs present the same discrepancy in time series length between Spotify and TikTok due to the reasons discussed in the previous section, namely the limited temporal coverage of TikTok data (Figure 19).

[20] https://www.bbc.co.uk/newsround/52164519



(a) “X (with 2 Chainz & Saudi)” by ScHoolboy Q



(b) “All I Want (From ‘High School Musical: The Musical: The Series’)” by Olivia Rodrigo

Fig. 19: Spotify and TikTok virality time series for songs with divergent model performances. To improve visual clarity, Spotify plots are truncated to 50 days after the end of the last wave.

In the case of “X (with 2 Chainz & Saudi)” (Figure 19a), the wave-based approach fits the Spotify time series with high accuracy but fails to capture any meaningful dynamics on TikTok. This failure is primarily explained by data sparsity: the TikTok Research API returns very few videos associated with the official audio of this song. In fact, the maximum number of daily videos using this audio is only two, resulting in an extremely sparse and noisy time series. This highlights the fact that relying solely on official audio IDs may not be sufficient for capture virality dynamics, as much of a song’s viral activity on TikTok may occur through unofficial or user-uploaded snippets that the Research API cannot consistently aggregate.

The opposite behavior happens for “All I Want (From ‘High School Musical: The Musical: The Series’)” (Figure 19b). In this case, the model provides a good fit for the TikTok time series, while the error associated with the Spotify series is higher. Despite the increased RMSE, visual inspection confirms that the model still captures the primary viral wave on Spotify at the start of the series. The error stems from a discrepancy between the fitted and observed speed of viral decay. In addition, another notable difference lies in the temporal positioning of the main viral wave across platforms. While the virality peak on Spotify occurs shortly after release, the TikTok peak appears approximately 300 days later, reflecting delayed adoption driven by platform-specific trends and user-generated content.

Overall, such outlier cases illustrate that RMSE alone is not an exhaustive measure of model suitability. Rather, visual analysis is essential to understand the qualitative nuances of how music spreads. Moreover, such an analysis evidences a key limitation of our wave-based approach: although it improves flexibility by decomposing virality into independent diffusion events, it remains sensitive to data sparsity, delayed adoption, and non-epidemic decay patterns.

9 Concluding Remarks

In this article, we explored the use of epidemic models to represent and forecast music popularity on online social platforms. By distinguishing virality from long-term success on Spotify, we first evaluated how classic epidemic models fit both curves. After discovering that such models are more suitable for virality, we proposed a novel wave-based modeling approach that effectively captures multiple bursts of popularity, i.e., independent periods in which a song gains attention. Our results show our epidemic approach can represent viral dynamics with high interpretability and forecast performance comparable to traditional time-series methods.

We also perform a case study applying the same methodology to data from TikTok, one of the main platforms driving the phenomenon of online virality. Using data collected via the platform's Research API, we build time series for a sample of the songs used in the Spotify analysis and replicate the same analyses to verify whether this methodology is also effective for other platforms. The results demonstrate that the virality dynamics observed on TikTok are different from those on Spotify. Nevertheless, the application of our wave-based approach produces good results that are close to the real data, offering interpretable parameters that provide valuable insights into the diffusion of these songs. Finally, the forecasting experiments reveal that our model's performance is also comparable to that of traditional ones, allowing the analysis of a song's viral behavior once its virality wave has begun.

Overall, our findings highlight the potential of contagion-based frameworks for understanding music consumption, with practical applications for online trend detection and marketing strategy development. Furthermore, the successful application of our methodology in different platforms also represents an important step toward generalizing the modeling of the viral phenomenon as a social contagion process.

Limitations and Future Work. The discrete nature and undisclosed calculation method of Spotify charts further limit our insights. In addition, our TikTok song sample is relatively small and may not be representative of broader platform dynamics, limiting the generalizability of our cross-platform comparison and conclusions. Such a sample should be expanded with more recent data. Furthermore, daily video counts may not perfectly capture TikTok virality, which metrics like views or listening history (currently unavailable via the API) could represent more accurately. Finally, our study relies on aggregate time-series data, which does not explicitly account for network-level structures or exogenous factors such as algorithmic recommendations or marketing campaigns. Moreover, our forecasting assumes static, pre-identified wave boundaries.

The starting boundary of a new wave is not known in real-time forecasting. Future work could explore more complex wave detection algorithms and also consider models in which subsequent waves aggregate the decay of preceding ones.

Declarations

Acknowledgements. This work was partially supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Fundação de Amparo a Pesquisa do Estado de Minas Gerais (FAPEMIG), and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). We also acknowledge the partial support of the Spoke 1 “FutureHPC & BigData” of ICSC - Centro Nazionale di Ricerca in High-Performance-Computing, Big Data and Quantum Computing, funded by European Union - NextGenerationEU.

Funding. Open access funding to be defined.

Conflict of interest. The authors declare no conflict of interest.

Ethics approval and consent to participate. Not applicable.

Data availability. Data sharing is restricted by the terms of use of the source platforms; therefore, the datasets are not available for public distribution.

Author contributions. G.P.O. conducted the experiments. All authors contributed to the experimental design, interpreted the results, and wrote the manuscript.

References

Addai E, Yousefi N, Agarwal N (2024) Utilizing fractional order epidemiological model to understand high and moderate toxicity spread on social media platforms. In: ASONAM, Lecture Notes in Computer Science, vol 15212. Springer, pp 298–308, https://doi.org/10.1007/978-3-031-78538-2_26

Baumann F, Arora N, Rahwan I, et al (2026) Dynamics of algorithmic content amplification on tiktok. EPJ Data Sci 15(1):18. https://doi.org/10.1140/EPJDS/S13688-026-00629-2

Biasioli M (2024) “It wasn’t our song anymore”: Molchat Doma, the death of the reader and the birth of the TikToker. IASPM Journal 14(1):151 – 171. https://doi.org/10.5429/2079-3871(2024)v14i1.10en

Bjørnstad ON, Shea K, Krzywinski M, et al (2020a) Modeling infectious epidemics. Nature methods 17(5):455–457. https://doi.org/10.1038/s41592-020-0822-z

Bjørnstad ON, Shea K, Krzywinski M, et al (2020b) The seirs model for infectious disease dynamics. Nature methods 17(6):557–559. https://doi.org/10.1038/s41592-020-0856-2

Castiglione A, Cozzolino G, Moscato F, et al (2021) Cognitive analysis in social networks for viral marketing. IEEE Trans Ind Informatics 17(9):6162–6169. https://doi.org/10.1109/TII.2020.3026013

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

Chou JS, Ochi M, Sakaki T, et al (2020) Constructive approach for early extraction of viral spreading social issues from twitter. In: WebSci. ACM, Online, pp 96–105, https://doi.org/10.1145/3394231.3397899

Compte DL, Klug D (2021) "It's Viral!" - A Study of the Behaviors, Practices, and Motivations of TikTok Users and Social Activism. In: CSCW Companion. ACM, pp 108–111, https://doi.org/10.1145/3462204.3481741

Coulter A (2022) Marketing Agile Artists: How Music Labels Can Leverage TikTok's Virality. Journal of the Music and Entertainment Industry Educators Association 22(1):135–161. https://doi.org/10.25101/22.5

Cullen ZN, Lee AY, Davidson BM, et al (2025) Not just 'for you': How the algorithmic crystal mediates communication and identity work on tiktok's FYP. Proc ACM Hum Comput Interact 9(7):1–27. https://doi.org/10.1145/3757636

Duricic T, Kowald D, Schedl M, et al (2021) My friends also prefer diverse music: homophily and link prediction with user preferences for mainstream, novelty, and diversity in music. In: ASONAM. ACM, pp 447–454, https://doi.org/10.1145/3487351.3492706

Galante F, Ravazzi C, Vassio L, et al (2025) Dominance or fair play in social networks? A model of influencer popularity dynamics. In: ASONAM, Lecture Notes in Computer Science, vol 16323. Springer, pp 306–321, https://doi.org/10.1007/978-3-032-13821-7_26

Guerini M, Strapparava C, Özbal G (2011) Exploring text virality in social networks. In: ICWSM. The AAAI Press, pp 506–509, https://doi.org/10.1609/icwsm.v5i1.14169

Gurung MI, et al (2024) Are Narratives Contagious? Modeling Narrative Diffusion Using Epidemiological Theories. In: ASONAM, Lecture Notes in Computer Science, vol 15214. Springer, pp 303–318, https://doi.org/10.1007/978-3-031-78554-2_20

Kontopoulou VI, Panagopoulos AD, Kakkos I, et al (2023) A review of ARIMA vs. machine learning approaches for time series forecasting in data driven networks. Future Internet 15(8):255. https://doi.org/10.3390/fi15080255

Kumar S, Jha AN (2022) Fake News Goes Viral! Determination and Analysis of Virality of Socially Relevant Events in Digital Governance. In: ICEGOV. ACM, pp

376–380, https://doi.org/10.1145/3560107.3560165

Li Y, Shao L (2024) Using an epidemiological model to explore the interplay between sharing and advertising in viral videos. Scientific Reports 14(1). https://doi.org/10.1038/s41598-024-61814-9

Ling C, Blackburn J, Cristofaro ED, et al (2022) Slapping cats, bopping heads, and oreo shakes: Understanding indicators of virality in tiktok short videos. In: WebSci. ACM, pp 164–173, https://doi.org/10.1145/3501247.3531551

Mathew B, Dutt R, Goyal P, et al (2019) Spread of hate speech in online social media. In: WebSci. ACM, pp 173–182, https://doi.org/10.1145/3292522.3326034

Nika M, Wilding T, Fiems D, et al (2015) Going multi-viral: Synthedemic modelling of internet-based spreading phenomena. In: VALUETOOLS, p 50 – 57, https://doi.org/10.4108/icst.valuetools.2014.258221

Oliveira GP, Couto da Silva AP, Moro MM (2024a) A quantitative comparison of viral and hit songs in the Brazilian music market. Vórtex Music Journal 12:1–29. https://doi.org/10.33871/vortex.2024.12.8727

Oliveira GP, Couto da Silva AP, Moro MM (2024b) Analyzing the temporal relation between virality and success in the Brazilian music market. In: SEMISH. SBC, pp 157–168, https://doi.org/10.5753/semish.2024.2656

Oliveira GP, Vassio L, Couto da Silva AP, et al (2025) Contagious rhythms: A wave-based epidemic approach for music virality on social platforms. In: ASONAM, Lecture Notes in Computer Science, vol 16322. Springer, pp 182–196, https://doi.org/10.1007/978-3-032-13513-1_16

Ribeiro B (2014) Modeling and predicting the growth and death of membership-based websites. In: WWW, pp 653–664, https://doi.org/10.1145/2566486.2567984

Rizoiu M, Xie L, Sanner S, et al (2017) Expecting to be HIP: hawkes intensity processes for social media popularity. In: WWW. ACM, pp 735–744, https://doi.org/10.1145/3038912.3052650

Rizoiu M, Lee Y, Mishra S (2018) Hawkes processes for events in social media. In: Chang S (ed) Frontiers of Multimedia Research. ACM / Morgan & Claypool, p 191–218, https://doi.org/10.1145/3122865.3122874

Rosati DP, Woolhouse MH, Bolker BM, et al (2021) Modelling song popularity as a contagious process. Proceedings of the Royal Society A 477(2253):20210457. https://doi.org/10.1098/rspa.2021.0457

Sachak-Patwa R, Fadai NT, Van Gorder RA (2018) Understanding viral video dynamics through an epidemic modelling approach. Physica A: Statistical Mechanics and its Applications 502:416–435. https://doi.org/10.1016/j.physa.2018.02.083

Seufitelli DB, Oliveira GP, Silva MO, et al (2023) Hit song science: a comprehensive survey and research directions. J New Music Res 52(1):41–72. https://doi.org/10.1080/09298215.2023.2282999

Sharma RS, Morales-Arroyo M, Pandey T (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

Sivaraman NK, Baijal S, Muthiah SB (2023) On the usage of epidemiological models for information diffusion over twitter. Soc Netw Anal Min 13(1):133. https://doi.org/10.1007/S13278-023-01130-8

Taylor SJ, Letham B (2018) Forecasting at scale. The American Statistician 72(1):37–45. https://doi.org/10.1080/00031305.2017.1380080

Vassio L, Garetto M, Chiasserini C, et al (2022) Temporal dynamics of posts and user engagement of influencers on facebook and instagram. In: ASONAM. ACM, p 129–133, https://doi.org/10.1145/3487351.3488340