Contagious rhythms: A wave-based epidemic approach for music virality on social platforms

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

[1] Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
{gabrielpoliveira,ana.coutosilva,mirella}@dcc.ufmg.br
[2] Politecnico di Torino, Turin, Italy
luca.vassio@polito.it

Abstract. Social media and streaming platforms have reshaped music consumption, enabling songs to go viral and achieve commercial success. In this paper, we explore the use of epidemic models to represent, explain, and forecast music popularity on streaming platforms. We introduce a wave-based approach that captures multiple independent bursts of popularity, which is not possible using classic epidemic models. Using streaming data from Spotify for more than 1000 songs, we evaluate our approach's ability to fit and forecast virality over time, comparing its performance with traditional time-series forecasting methods. Our findings show that our approach effectively captures viral dynamics, whereas it is less adapted to other aspects of popularity, such as long-term success. Moreover, it offers forecast accuracy comparable to conventional time series algorithms, with the additional benefit of providing interpretable parameters that shed light on the underlying diffusion processes.

Keywords: epidemic models · music virality · time series analysis.

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 [7,13]. 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.[3] Furthermore, in 2024, such platforms accounted for 69% of the global recorded music revenue.[4]

The spread of songs across such digital platforms shares many similarities with the propagation of infectious diseases. For example, as diseases spread

[3] IFPI: https://www.ifpi.org/ifpis-global-study-finds-were-listening-to-more-music-in-more-ways-than-ever/
[4] IFPI Global Music Report: https://globalmusicreport.ifpi.org/

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 [15,19] or views [12,20].

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 [8], *success* typically refers to long-term commercial performance, which is usually measured by indicators including (but not limited to) sales and streams [21]. 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?*

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); and (iii) we evaluate the forecasting performance of our method against traditional time-series methods (Section 6). Overall, our results show that epidemic models effectively capture viral dynamics better than success, while achieving forecast performance comparable to conventional approaches.

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 [7,8,18] For example, the work of Ling et al. [13] reveals that virality on TikTok is related to factors beyond the users' followers, including the presence of text and point of view on videos. Indeed, social media has amplified virality by enabling fast and widespread content sharing. Understanding such dynamics has direct applications in marketing [4], fighting misinformation [11], and addressing other social issues [6,14].

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 [21]. However, music popularity can be understood as a broader concept with distinct facets, namely success and virality [16,17]. 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 [22] and diffusion processes [19].

Following the work of Centola and Macy [5], 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 [2,3]. These models have been adapted to digital contexts, from the diffusion of online narratives and information [9,23] to the spread of toxicity in online platforms [1]. In the music context, epidemic approaches have been used to model the temporal dynamics of song popularity by analyzing song downloads [15,19] and video views [12,20].

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 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 [17], 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

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

[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.

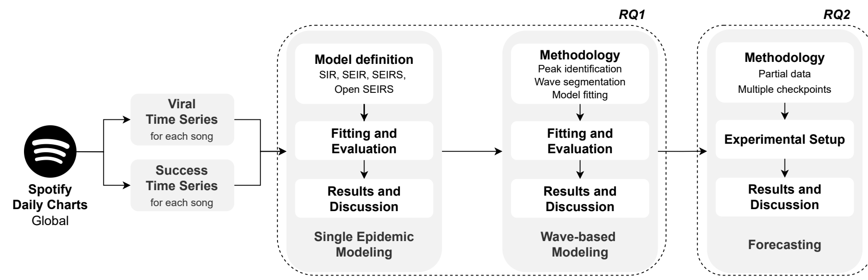

Fig. 1. Overview of the analysis conducted using our data and time series modeling.

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 [20], 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 [12,19], 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.

4.1 Model Definition

We consider four different epidemic models for our music popularity time series: SIR, SEIR, SEIRS, and Open SEIRS [2,3]. 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.

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

[8] Other normalization strategies are also possible, such as adding the noise first and then scaling by the maximum possible rank score.

a fixed population of $N = S+I+R$ and a closed epidemic, i.e., there are 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.

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.

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.

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.

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

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

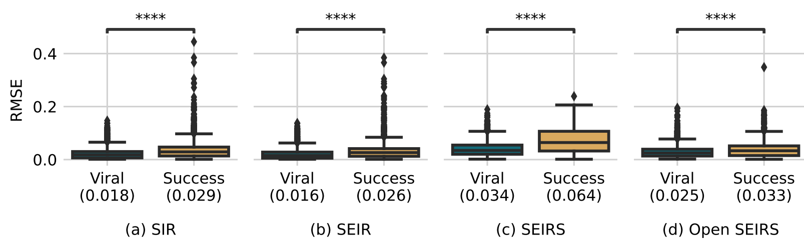

Fig. 2. RMSE for virality and success curves using single epidemic models. Values in parentheses are the median values. Significance is calculated using the Mann-Whitney U test: **** indicates $p \leq 0.0001$.

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 behavior by definition [8], 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 [20].

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.

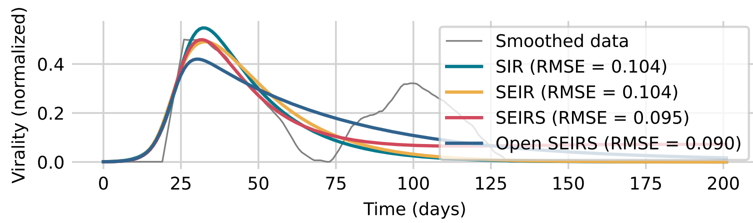

Fig. 3. Model fits for the virality of "Mon Amour - Remix" by Zzoilo and Aitana.

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.

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

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

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 [20].

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

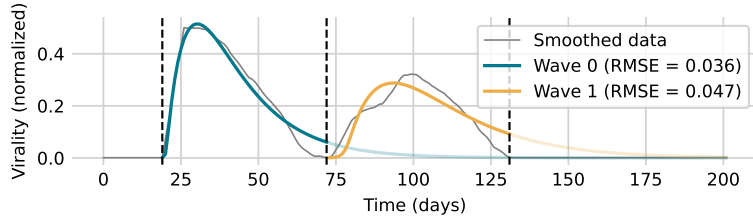

Fig. 4. Virality time series with the wave-based SEIR fit for the song "Mon Amour - Remix" by Zzoilo and Aitana. The vertical dashed lines delimit the waves.

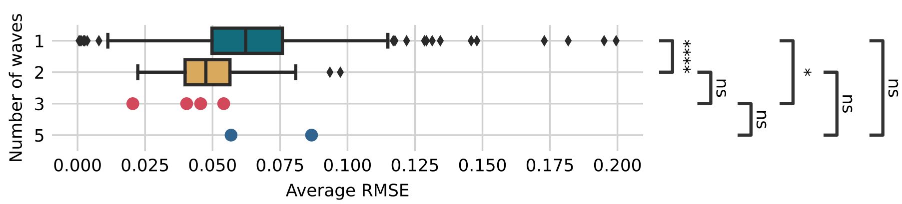

Fig. 5. Average RMSE distribution grouped by the number of identified waves. Significance is calculated using the Mann-Whitney U test: * for $0.01 < p \leq 0.05$; **** for $p \leq 0.0001$; and 'ns' for $p > 0.01$.

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 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 1 presents descriptive statistics for the primary SEIR parameters (i.e., β, γ, σ). Since a song may have multiple

Table 1. 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 2. 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

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 2 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 individuals 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 [16]. 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.

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 [18], 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. Unlike systems that may require dynamic detection of waves, here we assume the waves are already identified a-priori, based on the complete time series. This setup allows us to verify how well the SEIR model can forecast a song's future virality once a wave has already begun. In practical terms, this simulates 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.

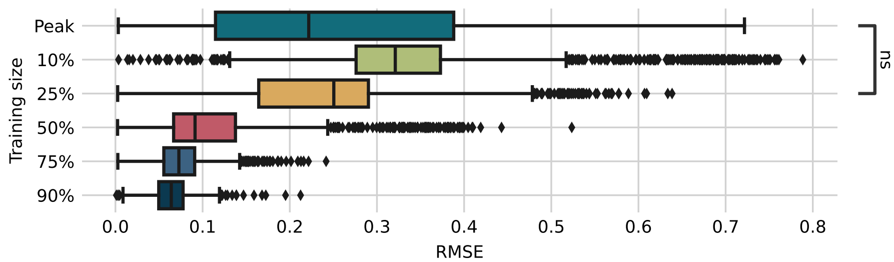

Fig. 6. Forecast RMSE for different partial data sizes. Unless specified with 'ns' ($p > 0.01$), all pairs of distributions are statistically different with $p \leq 0.0001$ (Mann-Whitney U test).

6.2 Experimental Setup

To evaluate the forecasting performance of the SEIR model, we perform two main experiments at the wave level. 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 [10], 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 [24]. For each baseline, we use the implementation provided by libraries *pmdarima*[11], *scikit-learn*[12], and *prophet*,[13] respectively, all with default parameter settings. 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 6, which presents the RMSE distribution for each fitting data size. As expected, increasing the amount of available data generally improves the forecasting performance, and all pairwise comparisons between training data sizes have statistically significant differences, except the pair between Peak and 25%. Indeed, the peaks in median occur at

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

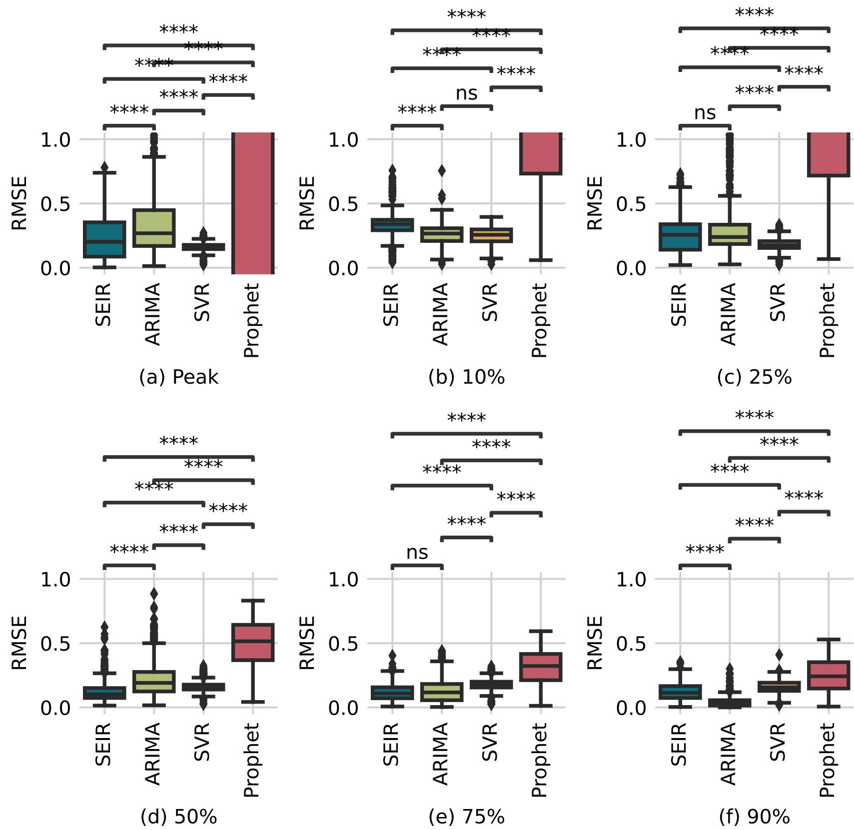

Fig. 7. Forecasting performance for our approach with baselines. Significance is calculated using the Mann-Whitney U test: **** for $p \leq 0.0001$.

33% of the wave time. In addition, using 50% of the wave to fit the SEIR model produces a median RMSE of 0.091, 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 548 songs for which all models (SEIR, ARIMA, SVR, and Prophet) successfully completed the forecast. Figure 7[14] reveals that no single model dominates across all training conditions. However, SEIR performs consistently well, outperforming others at the Peak and 50% marks. In contrast, ARIMA tends to be better when more data is available (especially at 90%), capturing the final gradual virality

[14] We set the upper limit of the y-axis to 1 to prevent distortions caused by extreme errors, particularly in Figures 7(a–c), in which Prophet failed to capture the underlying dynamics and produced significantly large forecasting errors.

decays at the end of waves. SVR and Prophet show competitive results in certain conditions, but they generally exhibit higher variability than SEIR.

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.

7 Concluding Remarks

In this paper, we explored the use of epidemic models to represent and forecast music popularity on streaming platforms. By distinguishing virality from long-term success, 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. 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.

Limitations and Future Work. Our analyses are limited to Spotify data and its Viral 50 charts, which may not reflect the full complexity of music virality on other platforms such as TikTok and Instagram. The discrete nature and undisclosed calculation method of such charts further limit our insights. 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 dynamic wave detection algorithms and accounting for their trade-offs.

Acknowledgments. This work was partially supported by CAPES, FAPEMIG, and CNPq (Brazil), and also by 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.

Disclosure of Interests. The authors do not have competing interests to declare.

References

1. Addai, E., et al.: 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, pp. 298–308. Springer (2024)
2. Bjørnstad, O.N., et al.: Modeling infectious epidemics. Nature methods **17**(5), 455–457 (2020)

3. Bjørnstad, O.N., et al.: The seirs model for infectious disease dynamics. Nature methods **17**(6), 557–559 (2020)
4. Castiglione, A., et al.: Cognitive analysis in social networks for viral marketing. IEEE Trans. Ind. Informatics **17**(9), 6162–6169 (2021)
5. Centola, D., Macy, M.: Complex contagions and the weakness of long ties. American Journal of Sociology **113**(3), 702–734 (2007)
6. Chou, J.S., et al.: Constructive approach for early extraction of viral spreading social issues from twitter. In: WebSci. pp. 96–105. ACM, Online (2020)
7. Compte, D.L., Klug, D.: "It's Viral!" - A Study of the Behaviors, Practices, and Motivations of TikTok Users and Social Activism. In: CSCW Companion. pp. 108–111. ACM (2021)
8. Guerini, M., et al.: Exploring text virality in social networks. In: ICWSM. pp. 506–509. The AAAI Press (2011)
9. Gurung, M.I., et al.: Are Narratives Contagious? Modeling Narrative Diffusion Using Epidemiological Theories. In: ASONAM. Lecture Notes in Computer Science, vol. 15214, pp. 303–318. Springer (2024)
10. Kontopoulou, V.I., et al.: A review of ARIMA vs. machine learning approaches for time series forecasting in data driven networks. Future Internet **15**(8), 255 (2023)
11. Kumar, S., Jha, A.N.: Fake News Goes Viral! Determination and Analysis of Virality of Socially Relevant Events in Digital Governance. In: ICEGOV. pp. 376–380. ACM (2022)
12. Li, Y., Shao, L.: Using an epidemiological model to explore the interplay between sharing and advertising in viral videos. Scientific Reports **14**(1) (2024)
13. Ling, C., et al.: Slapping cats, bopping heads, and oreo shakes: Understanding indicators of virality in tiktok short videos. In: WebSci. pp. 164–173. ACM (2022)
14. Mathew, B., et al.: Spread of hate speech in online social media. In: WebSci. pp. 173–182. ACM (2019)
15. Nika, M., et al.: Going multi-viral: Synthedemic modelling of internet-based spreading phenomena. In: VALUETOOLS. p. 50 – 57 (2015)
16. Oliveira, G.P., et al.: A quantitative comparison of viral and hit songs in the Brazilian music market. Vórtex Music Journal **12**, 1–29 (2024)
17. Oliveira, G.P., et al.: Analyzing the temporal relation between virality and success in the Brazilian music market. In: SEMISH. pp. 157–168. SBC (2024)
18. Ribeiro, B.: Modeling and predicting the growth and death of membership-based websites. In: WWW. pp. 653–664 (2014)
19. Rosati, D.P., et al.: Modelling song popularity as a contagious process. Proceedings of the Royal Society A **477**(2253), 20210457 (2021)
20. Sachak-Patwa, R., et al.: Understanding viral video dynamics through an epidemic modelling approach. Physica A: Statistical Mechanics and its Applications **502**, 416–435 (2018)
21. Seufitelli, D.B., et al.: Hit song science: a comprehensive survey and research directions. J. New Music Res. **52**(1), 41–72 (2023)
22. Sharma, R.S., et al.: 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 (2011)
23. Sivaraman, N.K., et al.: On the usage of epidemiological models for information diffusion over twitter. Soc. Netw. Anal. Min. **13**(1), 133 (2023)
24. Taylor, S.J., Letham, B.: Forecasting at scale. The American Statistician **72**(1), 37–45 (2018)