On the dynamics of music virality and its relation with mainstream success

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

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

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

Abstract. *Music going viral is not new, but the Web and social platforms have amplified its global reach. While related, virality and commercial success remain distinct. This work analyzes music virality and its relation to mainstream success from a data science perspective through three steps: characterization, causality analysis, and temporal modeling. Results show that artist-related and temporal features best distinguish both phenomena, which can sometimes predict each other. Furthermore, epidemiological models capture viral dynamics more accurately than long-term success. Overall, this work reinforces the differences and symbiotic links between virality and success, contributing to computing, musicology, and the music industry.*

Thesis Information

Category	PhD
Author	Gabriel Pereira de Oliveira
Graduate Program	Programa de Pós-Graduação em Ciência da Computação (PPGCC)
University	Universidade Federal de Minas Gerais (UFMG)
Advisor	Prof. Mirella M. Moro, PhD
Co-advisor	Prof. Ana Paula Couto da Silva, PhD
Doutorado Sanduíche	Politecnico di Torino (Italy), supervised by Luca Vassio and Marco Mellia
Defense and Approval Date	October 31st, 2025
Grade/Concept	Approved
Honors or Distinctions	Not applicable

Examination Committee

Member	**Affiliation**
Prof. Jussara M. Almeida, PhD	Departamento de Ciência da Computação – UFMG, Brazil
Prof. Flávio L. Schiavoni, PhD	Departamento de Ciência da Computação – UFSJ, Brazil
Prof. Luca Vassio, PhD	Dipartimento di Automatica e Informatica – Politecnico di Torino, Italy
Prof. Lucas N. Ferreira, PhD	Departamento de Ciência da Computação – UFMG, Brazil

Highlights

- Music virality and success are considered distinct yet interconnected phenomena;
- Artist-related and temporal features strongly discriminate virals from hits;
- Causality analysis reveals that there is no universal virality-success pathway;
- Novel wave-based epidemic model that captures multiple bursts of music virality;
- Epidemic approach generalizes across platforms: validated on both Spotify and TikTok.

1. Context and Problem

Every day, people have access to a massive volume of content on the Web, especially on Social Networks. In social media, "going viral" means that specific content spreads quickly across social platforms, being shared by thousands or even millions of users in a very short time span [Guerini et al. 2011]. Music is no exception, with viral songs widely shared in a short amount of time, and they may (but not necessarily) become successful by reaching the top of the charts with millions of streams and digital sales. Despite this, virality and commercial success remain conceptually distinct: a song can go viral without sustaining long-term chart performance, and vice versa. Understanding the interplay between these two phenomena is of paramount importance not only for the music industry but also for the broader fields of social computing and data science.

Therefore, in this thesis, we tackle the problem of understanding and modeling the music virality phenomenon on social platform data, as well as uncovering its relation with mainstream success. From a data management and social computing perspective, the challenge lies in (i) properly characterizing the two phenomena from heterogeneous data sources; (ii) uncovering the causal temporal dynamics between virality and success; and (iii) modeling the propagation process in a principled, interpretable way. Existing literature largely treats virality and success as synonymous or focuses on either domain in isolation [Sachak-Patwa et al. 2018, Rosati et al. 2021, Li and Shao 2024], leaving a critical gap in understanding their symbiotic relationship.

2. Research Goal

This thesis aims to **understand the phenomenon of music viralization on social platforms and its relationship with mainstream success**. We do so by using a data-driven methodology, combining disciplines such as machine learning, time series analysis, and epidemiology-inspired modeling. Specifically, our analyses are divided into three specific research goals (RGs):

RG1. Characterize viral and hit songs to identify the factors that distinguish them;
RG2. Analyze music virality as a stepping stone for mainstream success, i.e., the temporal relationship between both concepts; and
RG3. Model music virality as a process of social contagion for a deeper understanding of the dynamics of music viralization on such platforms.

3. Contributions

C1. Dataset: MGD+, an enhanced dataset on musical success over time in global and regional markets with enhanced artist and genre collaboration information. [Zenodo]
C2. Literature Review: A systematic review of 23 peer-reviewed studies on music virality (2009–2024), analyzing approaches, mechanisms, platforms, and methodologies;
C3. Quantitative Characterization: A data-driven binary classification of viral vs. hit songs shows F1-Scores above 0.7, with SHAP-based feature importance revealing that artist-related features and temporal features are the strongest discriminators. Acoustic features alone are insufficient.
C4. Causality Analysis: Granger Causality and Causal Discovery over Spotify data confirm that virality may forecast success (and vice versa). However, this cannot be generalized, reinforcing the distinction between the two phenomena.

C5. Epidemic Modeling: We propose a novel wave-based approach for music popularity, which decomposes a song's streaming trajectory into multiple epidemic waves. Besides Spotify, our approach is validated on TikTok data, with forecasting performance comparable to or surpassing classical methods in several scenarios.

4. Advances over the state-of-the-art

Prior work on music virality is either qualitative [Edmond 2014, Kahl 2012] or proposes epidemiological models validated on downloads and views from platforms such as BitTorrent and YouTube [Nika et al. 2015, Sachak-Patwa et al. 2018, Rosati et al. 2021], with no study systematically examining virality in modern streaming environments nor explicitly separating it from mainstream success. This thesis advances the state of the art on three fronts. First, unlike [Barbieri and Bonchi 2014] and [Seufitelli et al. 2022], we introduce a quantitative, multi-dimensional characterization that operationally distinguishes viral from hit songs using acoustic, lyrics-based, and extrinsic features from Spotify and Genius. Second, different from [Wallsten 2010], who investigates Granger causality only in a political YouTube context, we bring causal analysis to the streaming era by combining Granger Causality and Causal Discovery on global chart time series. Third, while [Nika et al. 2015] and [Sachak-Patwa et al. 2018] apply compartmental models to popularity data, neither addresses streaming charts or proposes a wave-based decomposition for multi-peak virality curves—a novel contribution of this thesis. Together, such advances constitute a novel integrated data science framework that combines characterization, causal analysis, and contagion modeling for music virality on contemporary streaming platforms.

5. Summary of the Solution

This thesis investigates music virality and its relationship with mainstream success through a three-stage data science framework applied to Spotify and TikTok chart data. **Quantitative characterization (RG1).** We integrate the MGD+ dataset with Spotify chart data (2017–2022) with Genius metadata and lyrics, then perform a comparative analysis of acoustic, intrinsic, and extrinsic features (including LIWC psycholinguistics and LDA topic modeling), followed by a classification experiment to distinguish viral from hit songs. Feature importance is then analyzed using SHAP values. **Causal analysis (RG2).** We model each song's chart performance into two time series (i.e., virality and success) and applied Granger Causality alongside PCMCI-based causal discovery to infer causal relationships between virality and success over time and verify whether the virality–success relationship can be generalized as a universal temporal precursor. **Epidemiological modeling (RG3).** We fit SIR, SEIR, SEIRS, and Open SEIRS compartmental models to streaming time series and propose a novel wave-based approach that segments each series into independent epidemic bursts, fitting a separate SEIR model per wave to capture multi-peak virality dynamics and enable competitive forecasting against ARIMA, SVR, and Prophet, which was validated on both Spotify and TikTok data.

6. Evaluation

Quantitative characterization (RG1). The classification experiment with 5-fold cross-validation shows that considering both intrinsic and extrinsic features yields the best F1-scores for distinguishing virals from hits (0.755 ± 0.009), with SHAP analysis confirming

that extrinsic features such as time-to-chart-entry and artist genres are the most discriminative factors. **Causal analysis (RG2).** Granger Causality shows that 61.86% of songs have a bidirectional temporal relation (i.e., virality can forecast success and vice versa), while PCMCI causal discovery, which has stronger causal assumptions, reveals that there is no causal link in 52.46% of songs. Such findings suggest that there is no universal virality-to-success pathway, and such a relationship may be strongly influenced by external factors, including the diffusion process itself. **Epidemiological modeling (RG3).** Whereas SEIR achieves the best median RMSE for virality on Spotify among traditional models, our wave-based approach yields a more realistic representation of virality patterns by capturing multiple engagement periods. Our forecasting experiments show that 50% of wave data suffices to achieve near-full-fit accuracy (RMSE 0.091 vs. 0.061), performing comparably or better than tradicional time series baselines (i.e., ARIMA, SVR, and Prophet). **Case Study on TikTok.** Replicating the epidemic modeling on TikTok daily video time series for 104 songs (sampled from 10% of our Spotify dataset), our wave-based approach outperforms all single-model baselines and, in several scenarios, also surpasses traditional forecasting methods. These results highlight the strong potential for generalizing the epidemiological framework to social media platforms more broadly.

7. Scientific and Technical Production

Direct Publications

1. **Oliveira, G. P.**; Vassio, L.; Couto da Silva, A. P.; Moro, M. M. Epidemic Wave Dynamics of Music Virality on Online Social Platforms. **Social Network Analysis and Mining (SNAM)**, 2026. [DOI]
2. **Oliveira, G. P.**; Vassio, L.; Couto da Silva, A. P.; Moro, M. M. Contagious rhythms: A wave-based epidemic approach for music virality on social platforms. In: **Intl. Conf. on Adv. in Social Network Analysis and Mining (ASONAM)**, 2025. [DOI]
3. **Oliveira, G. P.**; Couto da Silva, A. P.; Moro, M. M. What makes a viral song? Unraveling music virality factors. In: **ACM Web Science Conference (WebSci)**, p. 181-190, 2024. [DOI].
4. **Oliveira, G. P.**; Couto da Silva, A. P.; Moro, M. M. On the Causal Relationship Between Music Virality and Success. **IEEE Access**, v. 14, p. 122782-122791, 2025. [DOI].
5. **Oliveira, G. P.**; Couto da Silva, A. P.; Moro, M. M. A quantitative comparison of viral and hit songs in the Brazilian music market. **Vórtex Music Journal**, v. 12, p. 1-29, 2024. [DOI].
6. **Oliveira, G. P.**; Couto da Silva, A. P.; Moro, M. M. Music virality on social platforms: A literature review. Vórtex Music Journal, v. 13, p. 1-32, 2025. [DOI].
7. **Oliveira, G. P.**; Couto da Silva, A. P.; Moro, M. M. Analyzing the temporal relation between virality and success in the Brazilian music market. In: **Seminário Integrado de Software e Hardware (SEMISH)**, p. 157-168, 2024. [DOI].
8. **Oliveira, G. P.**; Vassio, L.; Couto da Silva, A. P.; Moro, M. M. Modeling music popularity as an epidemic: insights from the Brazilian market. In: **Brazilian Workshop on Social Network Analysis and Mining (BraSNAM)**, p. 79-92, 2025. [DOI].

Dataset

9. Seufitelli, D. B.; **Oliveira, G. P.**; Silva, M. O.; Moro, M. M. MGD+: An Enhanced Music Genre Dataset with Success-based Networks. In: **Dataset Showcase Workshop (DSW@SBBD)**, 2023. [DOI].

Publications Derived from Undergraduate Co-advisorships

10. Silva, B. R. A; **Oliveira, G. P.**; et al. Relações entre Clima, Vetores e Dengue no Espírito Santo: Uma Abordagem Baseada em Regressão, Correlação e Modelagem Epidemiológica. In: **Simp. Bras. de Computação Aplicada à Saúde (SBCAS)**, 2026. DOI.
11. Fonseca, G. A.; **Oliveira, G. P.**; Couto da Silva, A. P. Influência Externa na Popularidade Musical: Google Trends como Indicador de Viralidade e Sucesso no Spotify. In: **Braz. Works. on Social Network Analysis and Mining (BraSNAM)**, 2025. [DOI].

Publications on Other Music-related Aspects

12. **Oliveira, G. P.**; Moro, M. M. Frequent Genre Mining on Hit and Viral Songs. **J. of Information and Data Management (JIDM)**, v. 16, p. 136-145, 2025. [DOI].
13. Silva, M. O.; **Oliveira, G. P.**; Seufitelli, D. B.; Moro, M. M. Collaboration as a Driving Factor for Hit Song Classification. In: **Brazilian Symposium on Web and Multimedia (WebMedia)**, 2022. [DOI].
14. Silva, M. O.; **Oliveira, G. P.**; Moro, M. M. Data Insights on Gender Representation: Analyzing the Book and Music Industries. In: **Data Science for Social Good Brazilian Workshop (DS4SG@SBBD)**, 2024. [DOI].
15. Melo-Gomes, L.; Seufitelli, D. B.; **Oliveira, G. P.**; Silva, Mariana O.; Moro, M. M. Análise do Sucesso Musical no Brasil Utilizando Dados do Twitter. In: **Workshop de Trabalhos de Alunos de Graduação (WTAG@SBBD)**. 2022 [DOI].

Distinctions and Awards. (i) Best Paper Award at SEMISH 2024; (ii) Best Paper Runner-up at DSW 2023 and WebMedia 2022; (iii) CAPES/PDSE Scholarship – *Doutorado Sanduíche* at Politecnico di Torino, Italy (October 2024 to March 2025).

References

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

Edmond, M. (2014). Here we go again: Music videos after youtube. *Television and New Media*, 15(4):305 – 320.

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

Kahl, C. (2012). Create attention to attract attention - viral marketing of digital music in social networks. In *AMCIS*. Association for Information Systems.

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

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

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

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

Seufitelli, D. B. et al. (2022). From Compact Discs to Streaming: A Comparison of Eras within the Brazilian Market. *Revista Vórtex*, 10(1):1–28.

Wallsten, K. (2010). "Yes we can": How online viewership, blog discussion, campaign statements, and mainstream media coverage produced a viral video phenomenon. *Journal of Information Technology and Politics*, 7(2-3):163 – 181.