

Análise do Sucesso Musical no Brasil Utilizando Dados do Twitter

**Luiza de Melo-Gomes, Danilo B. Seufitelli, Gabriel P. Oliveira,
Mariana O. Silva, Mirella M. Moro**

Universidade Federal de Minas Gerais (UFMG)
Av. Antônio Carlos, 6627 – 31270-901 Belo Horizonte, MG

{luizademelo, daniloboechoat, gabrielpoliveira}@dcc.ufmg.br
{mariana.santos, mirella}@dcc.ufmg.br

Abstract. *Our goal is to analyze how Twitter data relates to the success of musical artistic careers. First, we collect data on tweets, number of likes and retweets for each artist profile from Spotify Charts in the Brazil. From the data collected, we build time series to represent the career of each artist, and then we investigate whether the most successful periods occur close to each other. Such exploratory analysis helps to identify temporal patterns that can reveal the existence of hot streaks, i.e., periods of above-normal success. Finally, we analyze the most frequent terms before and after the artists' peaks of success.*

Resumo. *Este estudo tem como principal objetivo analisar como os dados do Twitter se relacionam com o sucesso da carreira artística musical. Primeiro, realizamos a coleta de dados de tweets, número de curtidas e retweets para cada perfil de artista presente nas paradas de sucesso do Spotify no mercado musical brasileiro. A partir dos dados coletados, construímos séries temporais para representar a carreira de cada artista, e investigamos se os períodos de maior sucesso ocorrem próximos uns dos outros. Tal análise exploratória auxilia na identificação de padrões temporais que podem revelar a existência de hot streaks, i.e., períodos de sucesso acima do normal. Então, analisamos quais são os termos mais frequentes antes e depois dos picos de sucesso de artistas.*

1. Introdução

A indústria fonográfica enfrenta desafios diários devido à constante evolução de dados do universo musical. Boa parte da responsabilidade deste crescente volume de dados vem da internet, seja em plataformas como Billboard e Spotify ou até mesmo de redes sociais tais como o Twitter, Facebook e Instagram [Çimen and Kayis 2021, Seufitelli et al. 2022]. Em geral, tais plataformas fornecem dados musicais complexos, criando novas oportunidades de extração de conhecimento, beneficiando diferentes segmentos musicais. Para artistas a A&R (artistas & repertório) – responsáveis por encontrar e desenvolver talentos musicais – alcançar novos fãs, descobrir tendências e identificar novos talentos se tornaram tarefas não triviais que exigem ajuda automatizada [Garimella and West 2019].

Especificamente, o Twitter é uma das maiores redes sociais da atualidade, contando com mais de 330 milhões de usuários ativos mundialmente.¹ Desde sua criação,

¹<https://www.nytimes.com/2022/07/14/business/twitter-outage.html>

em março de 2006, o Twitter se transformou em uma grande fonte de divulgação de trabalhos artísticos. Nele, pessoas podem acompanhar o conteúdo gerado por artistas de preferência em textos curtos e objetivos (i.e., *tweets*). Além disso, conseguem interagir com tal conteúdo de diferentes formas, incluindo curtidas, *retweets* (i.e., compartilhamento), comentários, entre outras. Dessa forma, quanto mais interações recebidas, maior é o alcance do conteúdo de artistas, aumentando a probabilidade de expandir a base de fãs [Oliveira et al. 2020, Rousidis et al. 2020]. Nesse contexto, a dimensão social é uma das facetas do sucesso, explorar o engajamento do conteúdo criado por artistas nas plataformas sociais é essencial para compreender sua popularidade.

Este trabalho apresenta uma análise exploratória do engajamento de artistas musicais no Twitter. Em particular, definem-se séries temporais de artistas que apareceram nas paradas brasileiras de sucesso no Spotify. Tais séries são construídas a partir do engajamento dos *tweets* das contas de cada artista e representam sua popularidade nessa plataforma. Ao agrupar os períodos de maior sucesso na carreira de artistas, é possível identificar padrões temporais que podem revelar a existência de *Hot Streaks*. Este conceito pode ser definido como um período de sucesso acima da média, sendo estudado em diversos contextos profissionais, e.g., carreiras artísticas, culturais e científicas [Liu et al. 2018]. Os *insights* obtidos podem beneficiar diferentes setores na indústria da música, auxiliando no gerenciamento de carreiras e no planejamento de futuros lançamentos.

O restante deste artigo está estruturado da seguinte maneira: na Seção 2, são apresentados os trabalhos relacionados, necessários para contextualizar e permitir o entendimento do presente estudo. Na Seção 3, é detalhada a metodologia adotada para coletar o conjunto de dados. Na Seção 4, são apresentados os resultados da análise exploratória conduzida. Por fim, na Seção 5 apresentamos a conclusão e as propostas de continuação deste trabalho e destacamos as limitações.

2. Trabalhos Relacionados

A música sempre teve uma dimensão social, seja moldada pelo engajamento de artistas ou ouvintes. Diferentes serviços de mídia social (como Facebook, YouTube, Twitter) são projetados para atrair o público e incentivá-lo a descobrir novos artistas, compartilhar recomendações e consumir música. Ao longo dos anos, o número de estudos voltados para descobrir a receita para o sucesso musical aumentou consideravelmente, dando origem à área de *Hit Song Science* (HSS), que aborda o problema de prever a popularidade de músicas [Pachet 2011]. Nesse sentido, vários estudos analisam o impacto de diferentes fatores no sucesso musical, incluindo características acústicas [Martín-Gutiérrez et al. 2020, Vötter et al. 2021], letras [Trindade et al. 2021], colaboração [Silva et al. 2019, Oliveira et al. 2020], ou até mesmo engajamento em plataformas sociais [Calefato et al. 2018, Cosimato et al. 2019].

De fato, considerar métricas de engajamento social pode ser uma ferramenta valiosa ao medir o sucesso musical. Por exemplo, Abel et al. [2010] caracterizam álbuns de artistas a partir de informações retiradas de blogs. Ainda, Dewan e Ramaprasad [2014] estudam a relação entre a repercussão de músicas em blogs (*buzz*), a reprodução em rádios e vendas. Os resultados mostram que a relação entre o *buzz* da música e suas vendas é mais forte para músicas de nicho do que músicas *mainstream*. De uma perspectiva social diferente, Araujo et al. [2017] coletam mensagens do Twitter referentes a 30 dias anteriores

ao lançamento dos álbuns examinados. Também usando redes sociais, Ren e Kauffman [2017] consideram comentários sobre música no Last.fm após sua data de lançamento.

Apesar de reforçarem a importância de fatores extraídos de redes sociais na análise de sucesso musical, a maioria dos estudos mencionados se limitam a utilizar apenas menções e conteúdo dos ouvintes. Como diferencial, este trabalho busca explorar a importância do conteúdo criado pelos próprios artistas para avaliar o engajamento e sucesso alcançado nas plataformas sociais. Como o marketing tem grande impacto no sucesso futuro das músicas, considerar plataformas e recursos sociais pode fornecer fortes indicações do potencial de sucesso musical.

3. Metodologia

A metodologia deste trabalho é dividida em três etapas: aquisição de dados da Web, construção de séries temporais e detecção do agrupamento do sucesso, como segue.

Aquisição de Dados. Para coletar os dados, construímos um *crawler* em Python utilizando a API do Twitter.² Como ponto de partida da coleta, nós utilizamos os charts do Spotify para o mercado Brasileiro, do período de janeiro de 2017 até abril de 2022. Nós coletamos as informações extraídas da *timeline* do usuário (@s) de cada artista presente nos charts que possuíam perfis ativos no Twitter. Ou seja, coletamos seus *tweets*, *retweets*, número de curtidas e o número de *retweets* por *tweet*. Porém, notou-se a presença de homônimos de artistas, ou seja, usuários com o mesmo nome que artistas. Para minimizar a chance de coletar perfis de não artistas, nós implementamos uma heurística para identificar e validar o perfil de artista: busca de usuários pelo nome do artista; seleção dos 20 primeiros resultados; e escolha do perfil com o maior número de seguidores. Afinal, há maior probabilidade de artistas possuírem mais seguidores que usuários não famosos. Ao término desse processo heurístico, nossa base de dados continha um total de 791 perfis de artistas. Mesmo que o perfil selecionado tenha poucos seguidores, resolvemos não descartá-lo. É importante destacar que nós coletamos os *tweets* das *timelines* de cada artista desde março de 2006 até junho de 2022. Como resultado, nosso conjunto de dados final possui 2.093.994 *tweets*.

Construção das Séries Temporais. A evolução do sucesso de artistas no Twitter pode ser medida de várias formas. Uma métrica válida é considerar o número de curtidas que uma postagem recebe, ou então o número de compartilhamentos que suas publicações recebem, dentre outras abordagens. Neste trabalho, nós construímos a série temporal de artistas a partir da soma da distribuição do número de *retweets* e *likes* de cada postagem que artista realizou por dia (no intervalo de 2006 a 2022). Em outras palavras, cada ponto da série temporal de artista representa o sucesso acumulado em um determinado dia, de acordo com suas publicações feitas no twitter (*tweets*). Desta forma, a medida de sucesso considerada é o número total de *likes* (soma do número de vezes que as postagens foram favoritadas no twitter) e o número de *retweets* (número de vezes que as postagens foram compartilhadas) por dia.

Análise de Proximidade do Sucesso. Para melhor compreender as características de sucesso, esta seção apresenta uma análise com o objetivo de identificar se os períodos

²API do Twitter: <https://developer.twitter.com/en/apps>

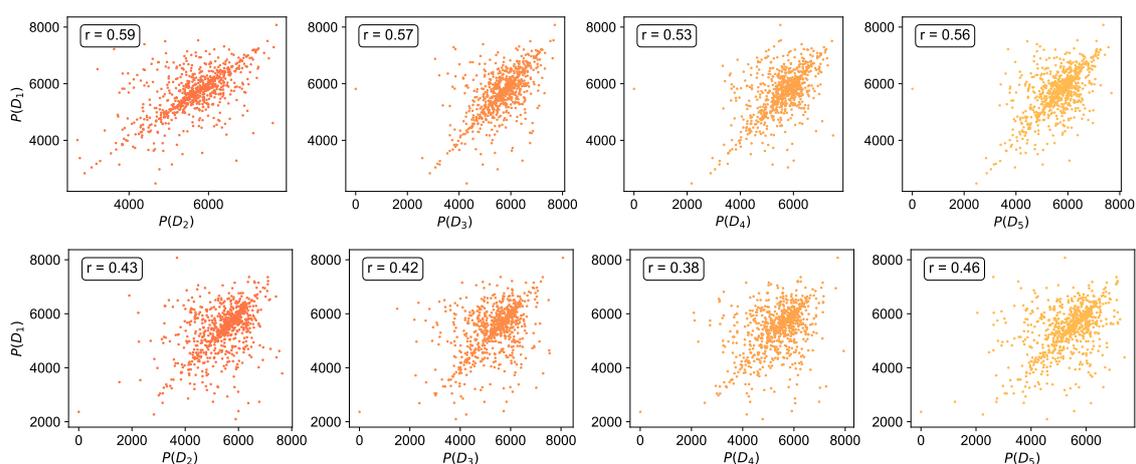


Figura 1. Gráficos de dispersão com correlação de Pearson (r) da posição do dia de maior sucesso nas carreiras dos artistas (D1) com D2, D3, D4 e D5, respectivamente. O sucesso é representado por likes (acima) ou retweets (abaixo). Cada ponto do gráfico representa um artista.

de sucesso de artistas ocorrem em intervalos temporais próximos uns dos outros. Desta forma, nossa análise está dividida em dois eixos principais. Primeiramente, nós verificamos se há uma correlação entre os dias de maior sucesso de acordo com o número de curtidas dos *tweets*. Em seguida, nós comparamos a carreira de artistas com carreiras aleatórias a fim de checar a robustez dos resultados. É importante destacar que carreiras aleatórias são carreiras geradas pelo computador com o intuito de realizar a comparação com carreiras reais e verificar as diferenças. Para validação e análise, recuperamos as cinco datas onde tivemos o maior sucesso e comparamos a primeira delas com as outras. A parte superior da Figura 1 mostra os gráficos de dispersão com a correlação de Pearson da posição da data com maior sucesso em número de curtidas relacionando-se com as outras quatro datas. A parte inferior da Figura 1 é semelhante porém relaciona os dias de acordo com o número de retweets que as postagens obtiveram no período. É possível inferir que os resultados demonstram uma relação linear para todas as carreiras artísticas.

4. Caracterização

Nesta seção nós descrevemos a base de dados por meio da investigação dos conteúdos dos *tweets* e da análise da distribuição do sucesso ao longo do tempo. Para melhor visualizar as características textuais dos *tweets*, nós construímos uma nuvem de palavras a partir de todo o texto recuperado das postagens, como mostra a Figura 2. Para sua construção, nós concatenamos o texto de todos os *tweets* e realizamos o pré-processamento deste texto com o auxílio da biblioteca NLTK em Python.³ Com o auxílio de tal biblioteca, nós removemos todas as *stop words*, sinais de pontuação e palavras com menos de duas letras para que a nuvem de palavras utilize apenas as palavras mais representativas possíveis. Além disso, também foram removidos quaisquer links para facilitar o processamento.

Para investigar o comportamento de postagens de artistas em relação aos períodos em que seus *tweets* fizeram maior sucesso, nós filtramos o conteúdo dos textos das postagens 30 dias antes e 30 dias após tal período de maior sucesso. Aqui, utilizamos como

³NLTK: <https://www.nltk.org/>



- No words to thank all the affection I received here.
- Sending love to everyone who is trying to heal from things they do not discuss
- @RadioX will be broadcasting our 2014 Finsbury Park show tomorrow night, Friday 8th May. Tune in from 8pm BST.
- Following our show in Rio, there are no further live shows planned. Please disregard fraudulent promotions.
- Check my TikTok in 30 min
- Listen to Light Switch on @Spotify's Today's Top Hits all tickets available today 10am local time

Figura 2. Nuvem de palavras (à esquerda) e exemplos de tweets (à direita)

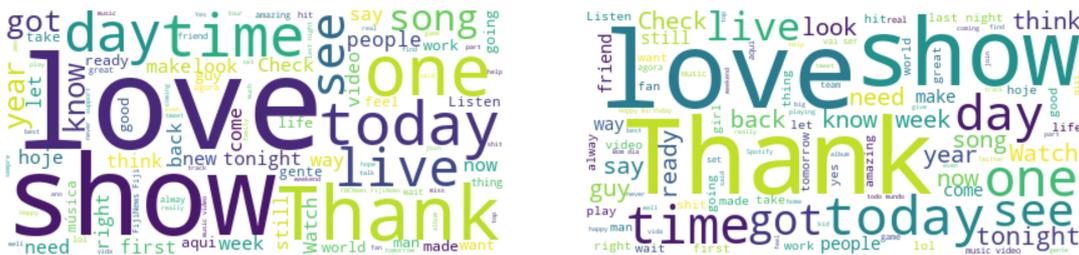


Figura 3. Nuvens de palavras extraídas dos textos dos *tweets* 30 dias antes (à esquerda) e 30 dia depois (à direita) do *tweet* de maior sucesso dos artistas.

medida de sucesso o número de curtidas. Em seguida, realizamos a comparação entre os *tweets* neste intervalo de 30 dias antes e depois. Como resultado (Figura 3), observamos que ambas as nuvens possuem muitas palavras em comum, embora com frequências distintas. Por exemplo, na nuvem à esquerda, as palavras *love*, *show* e *Thank* possuem diferentes tamanhos em relação à nuvem da direita. Apesar de tal dissimetria, notamos que não há diferença significativa entre o período anterior e posterior ao dia de maior sucesso. Em outras palavras, podemos dizer que artistas não mudam o teor de seus *tweets* após um *tweet* de muito sucesso.

Em seguida, investigamos a correlação entre a posição dos dias de maior sucesso de artistas de acordo com o número de curtidas e o número de compartilhamentos. A Figura 4 demonstra a relação temporal entre dias sucessivos. Note que há um leve declínio na correlação entre os dias ao longo do tempo. No entanto, esse padrão não é observado em carreiras aleatórias, nas quais a correlação é sempre entre -0.1 e 0.1 , o que sugere que há uma tendência de os períodos de maior sucesso nas carreiras de artistas ocorrerem próximas na perspectiva do engajamento social (*Twitter*). Por fim, calculamos a diferença entre os dois dias de maior sucesso de cada artista. Para isso, normalizamos essa diferença pelo número de dias N da série temporal de artistas. A Figura 5 mostra que essa distribuição tem o foco no ponto zero, o que sugere que esses dois dias estão próximos um do outro na série temporal. É possível notar que os resultados são similares tanto em relação ao número de curtidas, quanto ao número de *retweets*, o que vai de encontro com as análises anteriores. Além disso, ao compararmos com carreiras aleatórias, a distribuição torna-se bem diferente, o que demonstra que esse comportamento em carreiras musicais não é aleatório. Portanto, há evidência de que artistas experimentam momentos de sucesso fora do comum, ou Hot Streaks.

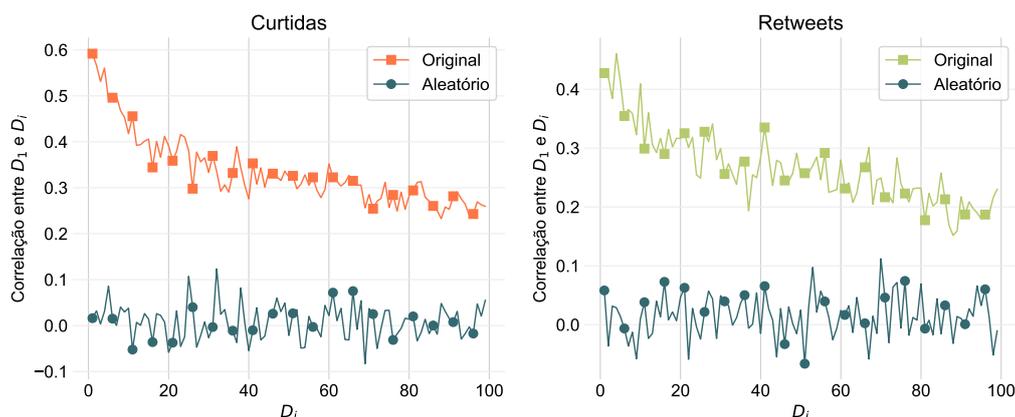


Figura 4. Correlação entre o primeiro e o dia de maior sucesso dos *tweets* dos artistas. À esquerda, a medida de sucesso adotada é o número de curtidas, e à direita o número de *retweets*.

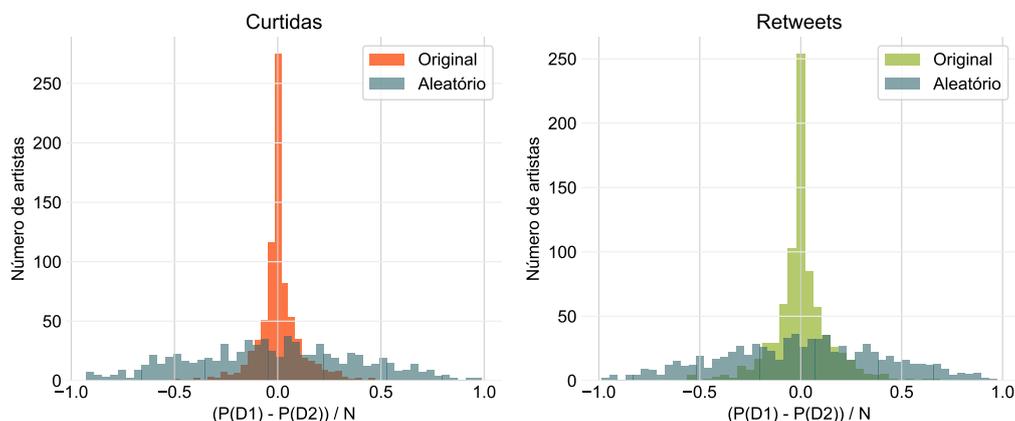


Figura 5. Diferença normalizada entre as posições do primeiro e segundo períodos de maior sucesso dos artistas. À esquerda, a medida de sucesso adotada é o número de curtidas, e à direita o número de *retweets*.

5. Conclusão e Trabalhos Futuros

Neste trabalho, analisamos a relação entre a interação social no Twitter e o sucesso de carreiras artísticas musicais. Considerando artistas presentes nas paradas brasileiras de sucesso no Spotify e informações coletadas de suas contas no Twitter, construímos séries temporais para representar suas carreiras artísticas ao longo do tempo. Através de uma análise exploratória, identificamos padrões temporais que sugerem a presença de *hot streaks*, i.e., períodos de sucesso acima da média. Tal resultado é relevante para a comunidade científica ao contribuir para o desenvolvimento de modelos mais complexos, enquanto para a indústria da música, ajuda a entender o comportamento dos ouvintes e as tendências de sucesso para artistas.

Trabalhos futuros envolvem resolver limitações do presente trabalho, incluindo o tratamento adequado do texto extraído dos *tweets*. Atualmente, os *tweets* coletados não foram separados de acordo com a linguagem, no momento da construção das nuvens de palavras. Portanto, pretende-se fazer esta verificação como melhoria do trabalho. Além disso, planeja-se realizar a detecção e caracterização de *hot streaks* no mercado musical brasileiro, avaliando em detalhes a evolução das séries temporais construídas.

Agradecimentos. Este trabalho foi financiado por CAPES, CNPq e FAPEMIG.

Referências

- Abel, F. et al. (2010). Analyzing the blogosphere for predicting the success of music and movie products. In *ASONAM*, pages 276–280, Odense, Denmark.
- Araujo, C. V. et al. (2017). Predicting music success based on users’ comments on online social networks. In *WebMedia*, pages 149–156, Brazil.
- Calefato, F. et al. (2018). Collaboration success factors in an online music community. In *ACM GROUP*, Sanibel Island, USA.
- Çimen, A. and Kayis, E. (2021). A longitudinal model for song popularity prediction. In *DATA*, pages 96–104. SciTePress.
- Cosimato, A. et al. (2019). The conundrum of success in music: Playing it or talking about it? *IEEE Access*, 7:123289–123298.
- Dewan, S. and Ramaprasad, J. (2014). Social media, traditional media, and music sales. *Mis Quarterly*, 38(1).
- Garimella, K. and West, R. (2019). Hot streaks on social media. In *International Conference on Web and Social Media*, pages 170–180. AAAI Press.
- Liu, L., Wang, Y., Sinatra, R., Giles, C. L., Song, C., and Wang, D. (2018). Hot streaks in artistic, cultural, and scientific careers. *Nature*, 559(7714):396–399.
- Martín-Gutiérrez, D. et al. (2020). A multimodal end-to-end deep learning architecture for music popularity prediction. *IEEE Access*, 8:39361–39374.
- Oliveira, G. P., Silva, M. O., Seufitelli, D. B., Lacerda, A., and Moro, M. M. (2020). Detecting collaboration profiles in success-based music genre networks. In *ISMIR*, pages 726–732.
- Pachet, F. (2011). Hit song science. In Tao Li, Mitsunori Ogihara, G. T., editor, *Music Data Mining*, chapter 10, pages 305–326. CRC Press, New York, NY, USA.
- Ren, J. and Kauffman, R. J. (2017). Understanding music track popularity in a social network. In *ECIS*, pages 374–388, Guimarães, Portugal. AIS.
- Rousidis, D., Koukaras, P., and Tjortjis, C. (2020). Social media prediction: a literature review. *Multimedia Tools and Applications*, 79(9):6279–6311.
- Seufitelli, D. B., Oliveira, G. P., Silva, M. O., Barbosa, G. R. G., Melo, B. C., Botelho, J. E., Melo-Gomes, L. d., and Moro, M. M. (2022). From compact discs to streaming: A comparison of eras within the brazilian market. *Revista Vórtex*, 10(1).
- Silva, M. O., Rocha, L. M., and Moro, M. M. (2019). Collaboration Profiles and Their Impact on Musical Success. In *ACM SAC*, pages 2070–2077, Limassol, Cyprus.
- Trindade, I. et al. (2021). Análise das letras das músicas brasileiras mais tocadas nas rádios das Últimas seis décadas. In *SBB D WTAG*, pages 1–7. SBC.
- Vötter, M. et al. (2021). Novel datasets for evaluating song popularity prediction tasks. In *IEEE International Symposium on Multimedia (ISM)*, pages 166–173. IEEE.