

Sentimentos em Cena: uma Análise dos Comentários em Trailers de Filmes da Netflix Brasil no YouTube

Matheus Rodrigues Alves¹, Mateus Barros Macedo², Jaqueline Ribeiro¹,
Livia Mancine¹, Cleon Xavier Pereira Júnior²

¹Instituto Federal Goiano (IF Goiano) – Campus Ceres – Ceres-GO – Brasil

²Instituto Federal Goiano (IF Goiano) – Campus Iporá – Iporá-GO – Brasil

{matheus.alves2, mateus.barros}@estudante.ifgoiano.edu.br,

{jaqueline.ribeiro, cleon.junior}@ifgoiano.edu.br,

livia.mancine@ifgoiano.edu.br

Abstract. *The production of audiovisual content is an essential part of the entertainment industry, and the Netflix platform stands out in this market. Trailers of Netflix movies posted on platforms like YouTube generate a large volume of interaction data, making manual analysis impossible. Additionally, studies analyzing interaction data are scarce in the Portuguese language. The aim of this study was to investigate interaction data on movie trailers on YouTube through the Netflix Brazil channel. Initial results revealed a peak in comments post-trailer release, which tends to decrease drastically over time, as well as strong correlations among interaction indices.*

Resumo. *A produção de conteúdo audiovisual é uma parte essencial da indústria do entretenimento, e a plataforma Netflix se destaca nesse mercado. Trailers de filmes da Netflix postados em plataformas como YouTube, geram um grande volume de dados de interação, impossibilitando uma análise manual. Além disso, trabalhos para analisar dados de interação são escassos na língua portuguesa. O objetivo desse trabalho foi investigar dados de interação em trailers de filmes no YouTube através do canal Netflix Brasil. Os resultados iniciais revelaram um pico de comentários pós-lançamento dos trailers, mas que tende a reduzir drasticamente com o tempo, assim como fortes correlações entre os índices de interação.*

1. Introdução

A produção de filmes e séries desempenha um papel fundamental no ramo do entretenimento. Neste cenário, uma das plataformas mais relevantes e populares é a Netflix¹. Esta plataforma está entre os maiores serviços de produção e *streaming* [Marchi and Ladeira 2023]. A Netflix oferece uma variedade de conteúdo sob demanda, incluindo filmes, além de investir fortemente em produções originais que atraem uma gama de telespectadores e críticos da área.

Postagens de trailers em plataformas como YouTube² é uma estratégia adotada pela Netflix para movimentar os telespectadores [Kappaun and Oliveira 2022]. Este

¹<https://www.netflix.com/>

²<https://www.youtube.com/>

público, por sua vez, mostra suas impressões, opiniões e expectativas quanto as produções a partir de curtidas e comentários [MacLeod et al. 2015]. Portanto, analisar essas interações, é uma estratégia para medir o impacto e a aceitação dos materiais da Netflix [Falico et al. 2021], além de permitir compreender as tendências das interações [Silva and Serrano 2023].

Diante do volume de dados de novas produções, bem como dos comentários feitos, fazer a análise desses dados manualmente se torna inviável [Coutinho and Malheiros 2020]. Assim, utilizar de técnicas de Inteligência Artificial (IA), como o Processamento de Linguagem Natural (PLN), possibilita lidar com grande volume de dados, otimizando o processo de análise. Uma análise possível diz respeito aos sentimentos (positivo, neutro ou negativo) contidos nos comentários [Silva and Serrano 2023]. Investigar esse grande volume de dados pode fornecer informações valiosas tanto para a criação de recomendações personalizadas de filmes, quanto para entender tendências de interação e aceitação do público. Porém, existe uma predominância de análises de textos em inglês, demonstrando assim uma falta de trabalhos na língua portuguesa [Santos et al. 2023].

Em [Silva and Serrano 2023], são realizadas comparações entre as avaliações e classificações de sentimentos realizadas em diferentes redes sociais como Instagram, Twitter e YouTube, mostrando diferenças entre técnicas de classificação de sentimentos, incluindo o uso da interface ChatGPT³ como eficiente para análise de sentimento. Ainda nesse sentido, os trabalhos a seguir apresentam contribuições com relação à análise de comentários em filmes.

O trabalho de [Coelho et al. 2017] apresentou uma pesquisa envolvendo a mineração de opiniões sobre filmes e livros, utilizando bases de dados de redes sociais diferentes, como Twitter e Amazon. Este trabalho concentrou-se principalmente na análise das discussões em torno dos sentimentos extraídos das obras cinematográficas e literárias. O trabalho de [Hu et al. 2023] desenvolveu uma análise de comentários postados em trailers em inglês, estabelecendo uma comparação entre os sentimentos presentes em diferentes redes sociais, como o YouTube e Facebook. Em [Chen et al. 2020], os autores analisaram sentimentos sobre produções da Netflix, e focaram especificamente em comentários do YouTube da série *La Casa de Papel*, os autores buscaram compreender as expectativas quanto aos comentários em trailers acerca da continuidade da série.

O objetivo deste trabalho foi fazer um estudo de análise temporal de comentários a partir dos sentimentos contidos nestes comentários, utilizando trailers de filmes em português, postados na plataforma YouTube pela Netflix Brasil⁴, com foco em conteúdo na língua portuguesa. A pesquisa também traz uma análise quanto a quantidade de sentimentos positivos, negativos e neutros destes comentários. Neste trabalho foi utilizado como ferramenta de classificação o GPT por meio da interface ChatGPT. De acordo com [Macedo et al. 2023, Silva and Serrano 2023], o uso do GPT como ferramenta de classificação se mostrou mais eficaz em relação a outros métodos tradicionais.

³<https://chat.openai.com>

⁴<https://www.youtube.com/@NetflixBrasil>

2. Metodologia

A fim de atingir o objetivo desta pesquisa, foram realizadas algumas etapas que permitiram a execução deste trabalho. Foi inicialmente realizada uma busca exploratória a fim de definir o escopo. A partir disso, foram selecionados vídeos postados pelo canal Netflix Brasil, que continham em seu título a palavra “trailer” e que tinham como idioma principal a língua portuguesa (note que não necessariamente são produções nacionais). Um dos parâmetros de limitação da busca foi o tempo, onde foram selecionados apenas vídeos que foram publicados entre os anos de 2019 a 2024. Este parâmetro foi estipulado com o intuito de delimitar o escopo da pesquisa e fornecer dados mais atuais.

Ao término da etapa de seleção, foi criada uma base com metadados de 5 trailers de filmes e 5 trailers de séries que continham o maior número de visualizações no canal da Netflix Brasil (Tabela 1). Ainda, foram descartados documentários e demais categorias de conteúdos audiovisuais. O número de vídeos foi escolhido também com o intuito de proporcionar uma melhor delimitação do escopo da pesquisa. As requisições dos vídeos e dos metadados foram feitas usando a *Application Programming Interface* (API) disponibilizada pelo YouTube. A partir da seleção dos vídeos, foram realizadas requisições na API, extraindo dados dos conteúdos nos vídeos como os números de comentários, número de visualizações, duração dos vídeos, número de curtidas, quanto os comentários em si. Estes dados foram considerados como as variáveis deste estudo.

Tabela 1. Vídeos selecionados e data de publicação

idVídeo	title	data_mes_ano
1	La Casa de Papel: Parte 3 — Trailer oficial — Netflix	03/06/2019
2	Pequenos Grandes Heróis, com Priyanka Chopra Jonas e Pedro Pascal — Trailer oficial — Netflix	04/12/2020
3	Round 6 — Trailer oficial — Netflix	02/09/2021
4	365 Dias: Hoje — Trailer oficial — Netflix	07/04/2022
5	Continência ao Amor — Trailer oficial — Netflix	03/08/2022
6	O Enfermeiro da Noite — Trailer oficial — Netflix	07/09/2022
7	Enola Holmes 2 — Trailer oficial: Parte 2 — Netflix	11/10/2022
8	The Crown — Temporada 5 Trailer oficial — Netflix	20/10/2022
9	ONE PIECE - Live Action — Trailer teaser oficial — Netflix	17/06/2023
10	ONE PIECE: A Série — Trailer oficial — Netflix	22/07/2023

Esse processo de coleta com a API, foi feito utilizando a linguagem de programação Python no Google Collaboratory⁵. Essa etapa resultou em um conjunto de 25.049 comentários retirados dos vídeos. A partir da coleta dessa base inicial de dados, foi retirada uma amostra do conjunto de dados. Essa amostra foi retirada considerando todos os meses a partir da publicação de cada vídeo, coletando então aproximadamente 10% dos comentários de forma aleatória, resultando em 2.496 comentários.

A partir da coleta da amostra do conjunto de dados, foi realizada a análise de sentimentos dos 2.496 comentários, usando GPT-3.5, com a interface ChatGPT. Para a classificação foi definido o seguinte prompt “Você é um classificador de texto, e está classificando os sentimentos contidos nesses textos entre positivo, negativo ou neutro. Classifique os textos abaixo levando em consideração esses sentimentos (positivo, negativo ou neutro) classificando entre -1 e 1 no formato de ponto flutuante. Onde -1 significa totalmente negativo e 1 completamente positivo. Retorne em formato de tabela”.

⁵<https://colab.google/>

A partir da classificação de ponto flutuante realizada pela análise de sentimento do GPT, foram definidos valores que representam sentimentos positivos, negativos e neutros. Definindo os comentários positivos como aqueles classificados entre 0.5 e 1, neutros aqueles classificados entre -0.5 e 0.5 e negativos aqueles classificados entre -0.5 e -1.

Uma das etapas de análise buscou identificar a correlação entre os dados de interação, como curtidas e comentários, e metadados do vídeo, como tempo de duração e engajamento, foi feito um mapa de correlações de Pearson, também chamado de mapa de calor (Figura 1). Também foram feitos gráficos de variação do número de comentários por sentimento pelo tempo (em meses) de toda a base de dados, buscando identificar padrões de ocorrência dos sentimentos (Figura 3).

3. Resultados e discussão

Um mapa de calor foi criado para entender o comportamento dos dados, conforme mostrado na Figura 1. Números negativos indicam uma correlação negativa, o que significa serem grandezas inversamente proporcionais, enquanto números positivos indicam correlações positivas, indicando grandezas diretamente proporcionais. Neste trabalho, foram feitas as seguintes interpretações das correlações: valores entre 0.7 e 1 indicam uma correlação forte, entre 0.5 e 0.7, uma correlação moderada; entre 0.3 e 0.5, uma correlação fraca; e valores entre 0 e 0.3 indicam uma correlação desprezível [Akoglu 2018].

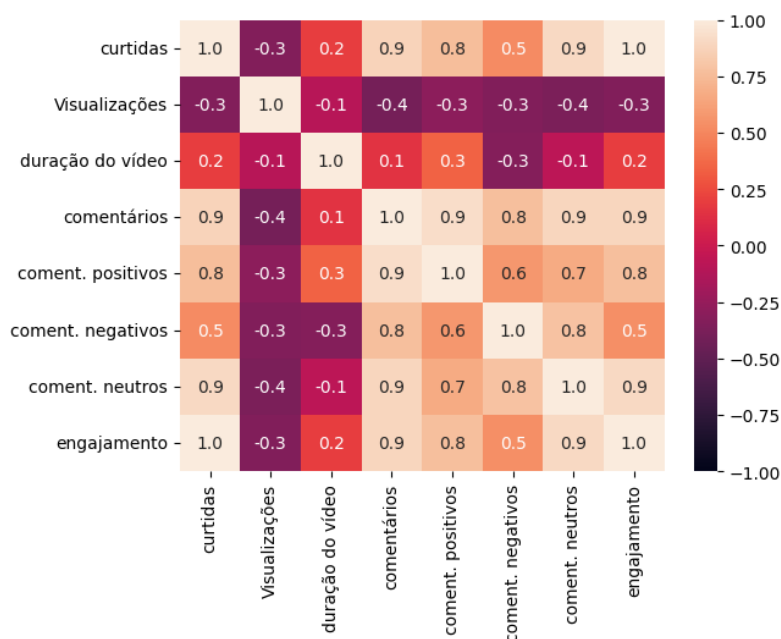


Figura 1. Mapa de Correlação

A partir do mapa de calor foi possível perceber que existe uma correlação negativa fraca entre o número de visualizações e as demais variáveis. Esse fator ocorre possivelmente pelo motivo de que com o passar do tempo e lançamento de novos títulos, mesmo que o número de visualizações aumente, a atenção e o engajamento dos telespectadores se voltam para os conteúdos mais recentes e em alta.

A medida de engajamento diz respeito a participação dos usuários em determinada publicação. Nesse caso o cálculo desse número foi feito com base na metodologia

proposta por [Silva et al. 2021], no qual o cálculo de engajamento é realizado a partir da média ponderada das métricas de interação. Com esse método os pesos para o cálculo são definidos considerando os valores das métricas.

Uma correlação apresentada no gráfico da Figura 1 diz respeito a quantidade de comentários positivos e o número de curtidas nos vídeos. O mapa de calor mostra que vídeos com comentários positivos tem uma correlação positiva forte com número de curtidas. Uma correlação positiva fraca presente no mapa (Figura 1), é com relação aos comentários positivos e a duração do vídeo, mostrando que vídeos maiores tem mais comentários classificados como positivo. O mesmo ocorre com índice de engajamento, que cresce com o tempo de duração do vídeo, podendo indicar que os telespectadores mostram mais comentários positivos e um maior engajamento com tempos de vídeos maiores, possivelmente por ser possível tem um panorama melhor do lançamento, como também proporcionar um maior tempo de interação com o conteúdo.

No tocante a análise de sentimentos, foram mostrados resultados condizentes com a literatura. O contingente de comentários classificados como “POSITIVO” obtiveram um percentual de 55.3%, apresentando a maior taxa de ocorrência. Comentários que se enquadraram na classe “NEGATIVO” representaram 16.9% e a ocorrências da classe “NEUTRO” representou 27.8%.

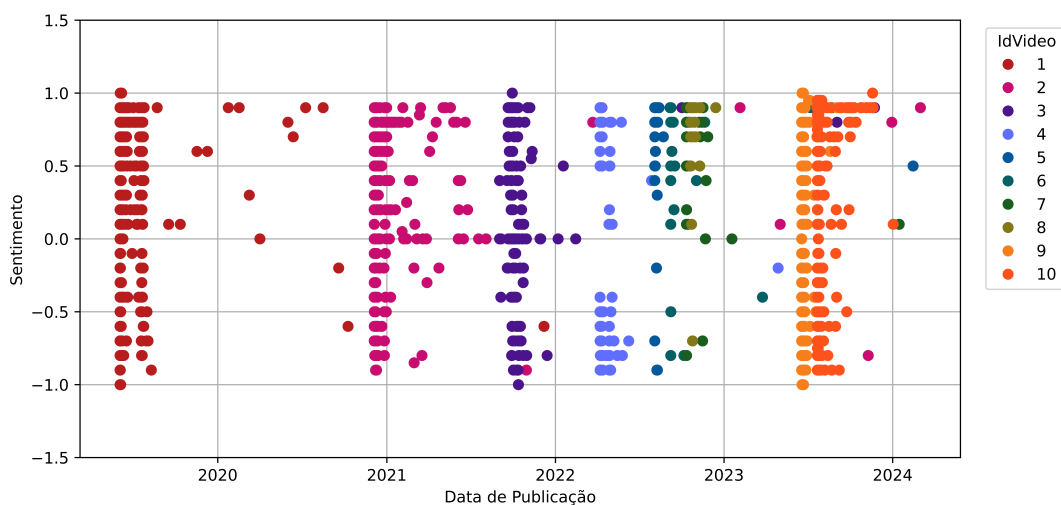


Figura 2. Ocorrência de sentimentos por tempo

A partir da tendência expressa na Figura 1 e também nos gráficos das Figuras 2 e 3, com o passar do tempo e número de visualizações a quantidade de interações diminui. Além disso, a intensidade dos sentimentos nos comentários é alterada, provavelmente devido à diminuição do engajamento e/ou comentários. A Figura 2, mostra a variação dos sentimentos nos comentários em uma escala temporal. É possível perceber que existem momentos onde ocorre uma concentração de comentários. Esses momentos são os dias de lançamento dos trailers, onde os usuários assistem e interagem mais. A figura 3, mostra a variação de sentimentos por tempo em cada vídeo, apresentando um comportamento padrão, onde o trailer é lançado, ocorre um pico de comentários, principalmente positivos, e com o tempo a quantidade de comentários cai para próximo de 0.

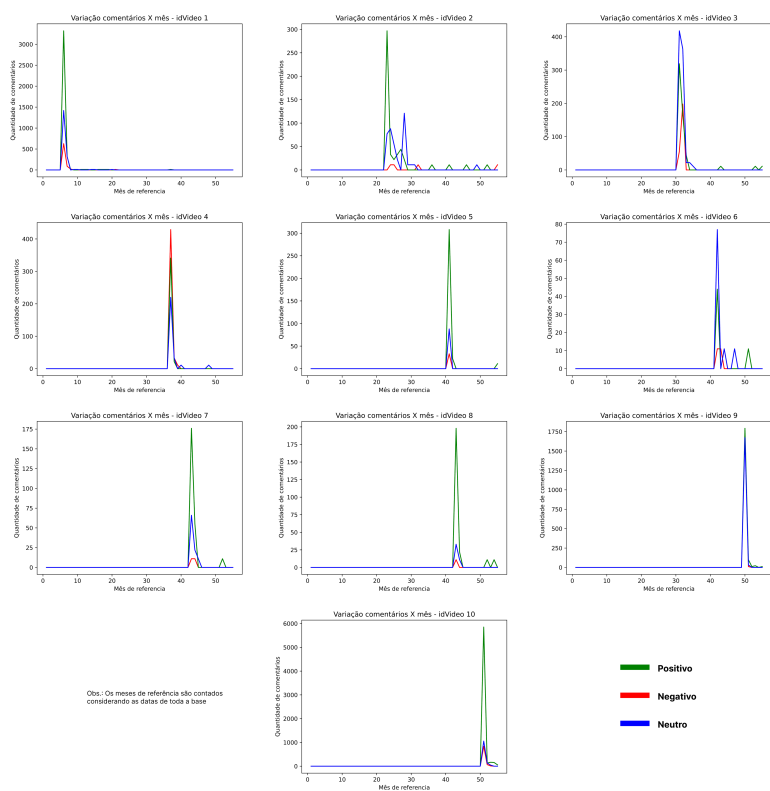


Figura 3. Ocorrência de sentimentos por tempo e por vídeo

Trailers como o de “O Enfermeiro da Noite” e “Pequenos Grandes Heróis” mostraram após seu lançamento, uma diferença sutil com relação à variação de comentários em comparação com outros trailers. Eles continuaram tendo ocorrência de comentários, o que não é comum nos outros vídeos. Com isso, os dados e análises são inconclusivos com relação a essa variação de comentários.

4. Considerações finais

O uso de técnicas para o processamento dos dados disponíveis em redes sociais, como o YouTube, se mostra um campo promissor por gerar metadados que permitem diferentes análises. Conhecimento a respeito das tendências dos usuários, bem como a recepção e engajamento dos mesmos com relação a um conteúdo, podem direcionar a produtora no planejamento e criação de conteúdos mais direcionados a um público. A exploração desses dados também podem colaborar com o público, oferecendo conteúdo personalizado. Como trabalhos futuros, fica a possibilidade de realizar comparações entre os sentimentos presentes em comentários de trailers de filmes e séries selecionados em redes sociais diferentes. Tal motivação ocorre uma vez que trabalhos como o de [Hu et al. 2023] apresentam pesquisas nesse sentido, mas em língua inglesa. É importante ressaltar que estes resultados estão delimitados e devem ser analisados e interpretados no contexto da pesquisa em andamento, de modo que uma base de dados diferente ou até mesmo a escolha de outros vídeos, podem apresentar resultados diferentes, reforçando a necessidade de continuidade da pesquisa.

Agradecimentos

As pessoas autoras deste trabalho agradecem o IF Goiano pelo apoio.

Referências

- Akoglu, H. (2018). User's guide to correlation coefficients. *Turkish journal of emergency medicine*, 18(3):91–93.
- Chen, R. N., Callista, A. S., Pratama, D. N., and Puspita, C. E. (2020). Sentiment analysis of youtube movie trailer comments using naïve bayes. In *Bulletin of Computer Science and Electrical Engineering*.
- Coelho, U. M., Lima, A. C. S., and O, N. (2017). Analisador de expressões positivas e negativas aplicado em comentários de livros e filmes.
- Coutinho, V. M. d. M. S. and Malheiros, Y. (2020). Detecção de mensagens homofóbicas em português no twitter usando análise de sentimentos. In *Anais do IX Brazilian Workshop on Social Network Analysis and Mining*, pages 1–12. SBC.
- Falico, L. L., Neves, S. C., and Romanelli, R. A. (2021). Balanço de produção: Mídia e educação: A influência do youtube no imaginário de crianças e adolescentes. In *GT 2 – Educação e Comunicação*. Mato Grosso, Brazil.
- Hu, Y., Chen, M., and Hui, S. (2023). Sentiment deviations in responses to movie trailers across social media platforms. In Springer, editor, *Marketing Letters: A Journal of Research in Marketing*.
- Kappaun, A. and Oliveira, J. (2022). Análise sobre viés de gênero no youtube: Um estudo sobre as eleições presidenciais de 2018 e 2022. In *Programa de Pós-Graduação em Informática (PPGI)*. Rio de Janeiro, RJ, Brazil.
- Macedo, M. B., Morais, B. G. S., de Almeida Ribeiro, L., da Costa, N. T., de Santana, T. S., and Junior, C. P. (2023). Análise comparativa de vídeos do youtube de canais de homens e mulheres na área de computação: uma investigação a partir dos comentários. In *Anais do XXXIV Simpósio Brasileiro de Informática na Educação*, pages 1862–1873. SBC.
- MacLeod, L., Storey, M.-A., and Bergen, A. (2015). Code, camera, action: How software developers document and share program knowledge using youtube. In *2015 IEEE 23rd International Conference on Program Comprehension*, pages 104–114.
- Marchi, L. and Ladeira, J. M. (2023). Originais netflix: Um panorama da produção audiovisual da netflix no brasil 2016-2019. In Escola de comunicação, a. e. d., editor, *REVISTA FAMECOS: mídia, cultura e tecnologia*.
- Santos, L. L., Bianchi, R. A., and Costa, A. H. (2023). Finbert-pt-br: Análise de sentimentos de textos em português do mercado financeiro. In *Anais do II Brazilian Workshop on Artificial Intelligence in Finance*, pages 144–155. SBC.
- Silva, I. O. d., Gouveia, F. C., et al. (2021). Engajamento informacional nas redes sociais: como calcular?
- Silva, M. E. R. and Serrano, P. H. S. M. (2023). Análise de sentimentos em textos de redes sociais: uma comparação entre o chatgpt e métodos tradicionais. *Cadernos de Comunicação*, 27(3).