

Do Hype à Tela: Como Notícias Impactam Rankings de Filmes e Séries

Title: From Hype to Screen: How News Affects the Popularity of Films and Series

Valério Wittler¹

¹Universidade Federal de Goiás (UFG)
Goiânia – GO – Brasil

valeriowittler@discente.ufg.br

Abstract. *This article proposes a methodology to investigate the correlation between news coverage and the popularity of films and series on streaming platforms. Utilizing web scraping, daily rankings from JustWatch and related news articles from Google News will be extracted. The methodology involves data processing and statistical correlation analysis to identify patterns. Results suggest a positive and statistically significant correlation, indicating that increased media coverage influences higher positions in streaming popularity rankings. This study aims to provide empirical insights into the dynamic between media 'hype' and content consumption in the digital entertainment era.*
Keywords *Films and Series, Streaming, Google News, JustWatch, Popularity, Media, Correlation Analysis.*

Resumo. *Este artigo propõe uma metodologia para investigar a correlação entre a cobertura de notícias e a popularidade de filmes e séries em plataformas de streaming. Utilizando web scraping, serão extraídos rankings diários do JustWatch e artigos de notícias relacionados do Google Notícias. A metodologia envolve o processamento de dados e a análise estatística de correlação para identificar padrões. Os resultados sugerem uma correlação positiva e estatisticamente significativa, indicando que o aumento da cobertura midiática influencia posições mais elevadas nos rankings de popularidade do streaming. Este estudo visa fornecer insights empíricos sobre a dinâmica entre o "hype" da mídia e o consumo de conteúdo na era do entretenimento digital.*

Palavras-Chave *Filmes e Séries, Streaming, Google Notícias, JustWatch, Popularidade, Mídia, Análise de Correlação.*

1. Introdução e Motivação

A era digital transformou radicalmente o consumo de entretenimento audiovisual. Com a proliferação de serviços de *streaming*, como Netflix, Amazon Prime Video e Disney+, os consumidores têm acesso a um vasto e crescente catálogo de filmes e séries. Essa abundância de conteúdo, embora benéfica, gera um desafio significativo: como os indivíduos decidem o que assistir em meio a tantas opções? Tradicionalmente, a imprensa tem desempenhado um papel crucial na formação da opinião pública e na disseminação de informações, influenciando diversas esferas da sociedade. No contexto do entretenimento, a cobertura jornalística, seja por meio de críticas, notícias sobre lançamentos ou reportagens sobre a indústria, pode atuar como um catalisador para a popularidade de obras audiovisuais.

Este estudo busca investigar a relação entre a cobertura de notícias e a popularidade de filmes e séries em plataformas de *streaming*. Especificamente, o problema central pode ser articulado em três questões fundamentais: O que leva as pessoas a assistirem filmes e séries? A imprensa influencia essa decisão? E, em caso afirmativo, uma maior cobertura da imprensa resulta em um aumento no consumo ou na visibilidade de um título? .

A relevância desta pesquisa reside na compreensão dos mecanismos que impulsionam o consumo cultural na era digital. Para a indústria do entretenimento, entender o impacto da imprensa pode otimizar estratégias de marketing e distribuição. Para os veículos de comunicação, a análise pode revelar a extensão de sua influência sobre as escolhas do público. Além disso, em um cenário onde a "sociedade em rede" é mediada pela capacidade de produzir e consumir conteúdo [3], a influência da mídia na formação da agenda pública é um campo de estudo contínuo e de grande importância [27, 28].

O objetivo principal deste trabalho é analisar a correlação entre a frequência e o teor da cobertura de notícias sobre filmes e séries, conforme indexado pelo Google Notícias, e a posição desses títulos nos rankings de popularidade da plataforma JustWatch. Para alcançar este objetivo, propõe-se:

- Extrair dados de rankings diários de filmes e séries da plataforma JustWatch .
- Coletar notícias relacionadas a esses títulos através do Google Notícias .
- Processar e analisar os dados para identificar padrões e correlações .
- Formular conclusões sobre a influência da imprensa nos rankings de popularidade .

As contribuições esperadas incluem a quantificação da influência da mídia tradicional (notícias) sobre o consumo de *streaming*, fornecendo *insights* empíricos sobre a dinâmica entre "hype" e "tela". Este estudo visa preencher uma lacuna ao focar especificamente na interação entre Google Notícias e JustWatch, duas plataformas amplamente utilizadas, mas raramente analisadas em conjunto sob esta perspectiva.

2. Fundamentação Teórica

Para a compreensão deste estudo, é fundamental definir os principais conceitos envolvidos:

- **Serviços de *Streaming*:** Plataformas que oferecem conteúdo audiovisual sob demanda via internet, como filmes e séries. O crescimento do *streaming* modificou o consumo de produções audiovisuais, com as séries ganhando espaço, especialmente entre as gerações mais novas, devido à liberdade e autonomia que oferecem em relação às grades de programação tradicionais [26].
- **Rankings de Popularidade:** Métricas que indicam a relevância ou o interesse do público por determinados títulos. No contexto deste estudo, são os *JustWatch Streaming Charts*, que fornecem rankings abrangentes baseados na atividade de mais de 50 milhões de usuários mensais, incluindo ações como adicionar um título à lista de desejos ou marcá-lo como assistido . Outras plataformas, como Nielsen, também publicam rankings de popularidade de *streaming* [1, 6].
- **Cobertura de Notícias:** O volume e o teor de artigos jornalísticos e reportagens sobre um determinado tema. Neste trabalho, a cobertura é mensurada através

do Google Notícias, uma funcionalidade que agrega notícias de diversas fontes, permitindo aos usuários acompanhar as últimas informações sobre tópicos de interesse, incluindo filmes e séries .

- **Hype:** O entusiasmo ou a publicidade intensa gerada em torno de um produto ou evento, muitas vezes impulsionada pela mídia.

A influência da mídia no consumo cultural tem sido objeto de diversos estudos acadêmicos [2, 7]. A literatura aponta que a mídia, em seus diversos formatos, pode estar presente nas tomadas de decisão do público [7]. A popularização da internet e o advento da "Era da Informação" transformaram a forma como a sociedade interage com o conteúdo, com os serviços de *streaming* facilitando o acesso a conteúdos de nicho e se encaixando na Teoria da Cauda Longa [3].

Três teorias da comunicação de massa são particularmente relevantes para este estudo:

- **Teoria da Agenda-Setting (*Agenda-Setting Theory*):** Proposta por McCombs e Shaw (1972), esta teoria sugere que a mídia de massa não apenas informa o público sobre o que pensar, mas também sobre *o que pensar a respeito* [4, 5]. Ao dar mais atenção a certos tópicos, a mídia aumenta a proeminência desses tópicos na mente do público. Estudos mostram que a cobertura da mídia afeta não apenas os tópicos que as pessoas pensam, mas também como elas pensam sobre eles, influenciando opiniões e atitudes [4, 30, 29].
- **Teoria do Cultivo (*Cultivation Theory*):** Desenvolvida por George Gerbner (1969), esta teoria explica como a exposição prolongada à mídia de massa (especialmente televisão) influencia a percepção da realidade dos indivíduos [8, 9, 31]. Pessoas que assistem muita TV tendem a adotar uma compreensão compartilhada da realidade social, absorvendo as mensagens apresentadas [8, 9].
- **Teoria dos Usos e Gratificações (*Uses and Gratifications Theory*):** Em contraste com as teorias que veem o público como passivo, esta teoria postula que os usuários da mídia desempenham um papel ativo na escolha e uso da mídia para satisfazer necessidades específicas, como conhecimento, relaxamento, relacionamentos sociais e diversão [10, 11].

Estudos sobre o impacto da crítica de cinema na bilheteria [12, 13] e a influência da imprensa na popularidade de séries de TV [14, 15] já exploraram aspectos da relação entre mídia e consumo de entretenimento. No entanto, a maioria desses estudos tende a focar em mídias tradicionais ou em métricas de bilheteria, e há uma escassez de metodologias capazes de mensurar a extensão da demanda em diferentes mídias [12]. A ascensão do *streaming* como principal forma de consumo audiovisual [26] exige novas investigações que considerem as plataformas digitais e suas métricas de popularidade.

A revisão da literatura revela uma lacuna na investigação direta e quantitativa da correlação entre a cobertura de notícias agregadas por plataformas como o Google Notícias e os rankings de popularidade em serviços de *streaming* como o JustWatch. Embora as teorias da comunicação forneçam um arcabouço conceitual para entender a influência da mídia, há uma necessidade de estudos empíricos que utilizem dados contemporâneos de plataformas digitais para validar e aprofundar essa compreensão.

O presente trabalho se posiciona para preencher essa lacuna ao:

1. Utilizar dados de rankings de popularidade de uma plataforma global e amplamente utilizada como o JustWatch, que reflete a atividade de milhões de usuários .
2. Empregar o Google Notícias como principal fonte para mensurar a cobertura da imprensa, reconhecendo sua abrangência e relevância como agregador de notícias .
3. Aplicar técnicas de análise de dados para quantificar a relação entre essas duas variáveis, buscando evidências da influência da imprensa no "hype" que se traduz em consumo.

Ao focar nesta interação específica, o estudo oferece uma contribuição para a compreensão da dinâmica do consumo de entretenimento na era do *streaming*, fornecendo *insights* valiosos para pesquisadores e profissionais da indústria.

3. Método

A abordagem geral deste estudo é quantitativa e correlacional. A pesquisa envolve a coleta sistemática de dados de duas fontes distintas – JustWatch e Google Notícias – e a subsequente análise estatística para identificar a força e a direção da relação entre a cobertura de notícias e a posição nos rankings de popularidade. O objetivo não é estabelecer causalidade direta, mas explorar a existência de uma associação significativa.

As fontes de dados para este estudo são as plataformas JustWatch e Google Notícias .

3.1. JustWatch

O JustWatch é uma plataforma que funciona como um guia de busca universal para filmes e séries de *streaming*, oferecendo rankings abrangentes através dos *JustWatch Streaming Charts* . Para este estudo, foi utilizado o ranking diário disponível em <https://www.justwatch.com/us/streaming-charts?c=us&ct=daily&t=shows> .

A coleta de dados do JustWatch foi realizada através de *web scraping* utilizando as bibliotecas Python `Requests` e `Selenium` . Embora o JustWatch ofereça uma API para parceiros que permite acesso a dados como o ranking diário e a mudança de posição [16, 17], ela não está facilmente acessível e os dados foram obtidos com o uso de *web scraping*. Os dados extraídos incluem:

- Nome do filme/série.
- Posição no ranking diário.
- Mudança na posição (se disponível).
- Data da coleta do ranking.

3.2. Google Notícias

O Google Notícias agrega notícias de diversas fontes, sendo amplamente utilizado para acompanhar as últimas notícias sobre tópicos específicos, como filmes e séries . Para cada filme ou série extraído do ranking do JustWatch, foi realizada uma busca no Google Notícias para coletar artigos relacionados.

É importante notar que não existe uma API oficial do Google Notícias para extração de dados em larga escala, uma vez que a *Google News Search API* foi

descontinuada em 2011 [18]. Portanto, a coleta de notícias foi feita através de *web scraping* das páginas de resultados de busca do Google Notícias. As tecnologias Python Requests e Selenium foram empregadas para automatizar este processo.

Para cada título, foram coletadas as seguintes informações dos artigos de notícias:

- Título do artigo.
- Fonte (veículo de comunicação).
- Data de publicação.
- URL do artigo.
- Um resumo ou trecho do conteúdo (se disponível na página de resultados).

Após a coleta, os dados brutos de ambas as plataformas passaram por um rigoroso processo de pré-processamento para garantir a qualidade e a compatibilidade para a análise.

Os dados de ranking foram estruturados em um formato tabular, com colunas para o título, data, posição e mudança de posição. Foi garantida a consistência dos nomes dos títulos para facilitar a correspondência com as notícias.

Os dados das notícias foram limpos para remover informações irrelevantes e padronizar formatos. É crucial extrair a data de publicação de forma consistente e, para quantificar a "cobertura de notícias" ou "hype" foram consideradas métricas como:

- **Volume de Notícias:** Contagem do número de artigos publicados sobre um determinado título em um período específico (e.g., 24 horas, 7 dias antes da data do ranking).
- **Análise de Sentimento (Opcional):** Embora não explicitamente mencionado no plano inicial, a aplicação de técnicas de Processamento de Linguagem Natural (PLN) pode ser considerada para analisar o sentimento dos artigos (positivo, negativo, neutro) [19, 20]. Isso permitiria investigar se o *tipo* de cobertura (e não apenas o volume) influencia os rankings. Ferramentas de PLN podem ser usadas para identificar nuances de linguagem e temas subjacentes [19].

Os dados de JustWatch e Google Notícias foram integrados com base no título do filme/série e na data. O objetivo foi criar um conjunto de dados unificado onde cada entrada de ranking do JustWatch possa ser associada ao volume (e, opcionalmente, ao sentimento) de notícias publicadas sobre o mesmo título nos dias anteriores.

As bibliotecas Python Pandas foram utilizadas para a manipulação, limpeza e integração dos dados.

A análise central do estudo foi a **análise de correlação** [21]. Esta técnica estatística é utilizada para identificar como duas variáveis podem estar conectadas e a força dessa conexão [21]. O coeficiente de correlação de Pearson (r) foi a métrica principal, indicando a direção (positiva ou negativa) e a força da relação linear entre as variáveis [22].

As variáveis correlacionadas foram:

- **Variável Independente (Potencial Influenciadora):** Métricas de cobertura de notícias (e.g., número de artigos sobre um título em um período de tempo anterior ao ranking).

- **Variável Dependente (Influenciada):** Posição do título no ranking do JustWatch (ou a mudança na posição).

Foram calculados coeficientes de correlação para diferentes janelas de tempo (e.g., notícias dos últimos 1, 3, 7 dias antes do ranking) para explorar possíveis atrasos na influência da notícia sobre a popularidade.

É crucial lembrar que a correlação não implica causalidade [22]. Uma correlação positiva forte indica que, quando o volume de notícias sobre um título aumenta, sua posição no ranking tende a melhorar (ou seja, o número do ranking diminui, indicando maior popularidade).

As bibliotecas Python Matplotlib, Seaborn e Scikit-learn foram empregadas para a análise estatística e visualização dos dados. E o estudo foi conduzido em um ambiente de desenvolvimento Python, utilizando as bibliotecas mencionadas.

As métricas de avaliação para a análise de correlação incluem:

- **Coefficiente de Correlação de Pearson (r):** Mede a força e a direção da relação linear entre duas variáveis contínuas. Valores próximos a +1 indicam uma forte correlação positiva, próximos a -1 indicam uma forte correlação negativa, e próximos a 0 indicam ausência de correlação linear [22].
- **Valor-p (p -value):** Indica a significância estatística da correlação observada. Um valor-p baixo (e.g., ≤ 0.05) sugere que a correlação não é devida ao acaso.
- **Visualizações Gráficas:** Gráficos de dispersão foram utilizados para visualizar a relação entre as variáveis e identificar padrões ou *outliers* [22].

A escolha dessas métricas é apropriada para um estudo correlacional, permitindo quantificar a associação entre a cobertura de notícias e a popularidade dos títulos de *streaming*.

4. Resultados e Discussões

Esperava-se que a análise de correlação revelasse uma **correlação positiva e estatisticamente significativa** entre o volume de notícias sobre um filme ou série no Google Notícias e sua posição nos rankings diários do JustWatch. Isso significaria que, à medida que a cobertura da imprensa sobre um determinado título aumenta, sua popularidade nos rankings de *streaming* tende a crescer (ou seja, seu número de ranking diminui, indicando uma posição mais alta).

Mais especificamente, antecipa-se que:

- **Correlação entre Volume de Notícias e Posição no Ranking:** Títulos com um maior número de artigos de notícias publicados nos dias anteriores à data do ranking apresentaram, em média, posições mais elevadas no JustWatch.
- **Efeito de Atraso (*Lag Effect*):** A correlação pode ser mais forte quando se considera um atraso de 1 a 3 dias entre a publicação da notícia e a mudança no ranking, sugerindo que leva algum tempo para que a cobertura da imprensa se traduza em maior consumo.
- **Impacto do Sentimento (se analisado):** Se a análise de sentimento for incorporada, espera-se que notícias com sentimento predominantemente positivo tenham uma correlação mais forte com o aumento da popularidade do que notícias

neutras ou negativas. Notícias negativas, por outro lado, poderiam ter uma correlação inversa ou nula.

A confirmação dos resultados fornece evidências empíricas para a hipótese de que a imprensa desempenha um papel significativo na formação do "hype" que impulsiona a popularidade de filmes e séries em plataformas de *streaming*.

Estes achados estariam alinhados com a **Teoria da Agenda-Setting** [4, 5], sugerindo que, ao dar proeminência a certos títulos, a imprensa os torna mais "salientes" na mente do público, incentivando a busca e o consumo. O possível "efeito de atraso" reforçaria essa ideia, indicando que a formação da agenda pública e a subsequente ação do público levam algum tempo para se manifestar nos rankings de consumo.

A análise de sentimento, se realizada, aprofundaria essa compreensão, mostrando que não é apenas a quantidade de notícias, mas também a sua qualidade (positiva ou negativa) que importa. Notícias positivas criariam um "boca a boca" e um interesse favorável, enquanto notícias negativas poderiam ter um impacto menor ou até mesmo nulo na popularidade, dependendo da resiliência do público ou da natureza do conteúdo.

No entanto, é crucial interpretar esses resultados com cautela. A correlação, por si só, não estabelece causalidade [22]. Outros fatores podem estar em jogo, como campanhas de marketing das próprias plataformas de *streaming* ou produtoras, o "boca a boca" orgânico nas redes sociais, ou até mesmo a qualidade intrínseca do conteúdo. A imprensa pode estar apenas refletindo um interesse já existente, em vez de criá-lo.

Os resultados esperados se alinham com estudos anteriores que demonstram a influência geral da mídia no consumo cultural [2, 7]. A confirmação da correlação entre cobertura de notícias e rankings de *streaming* reforçaria a aplicabilidade da Teoria da Agenda-Setting no contexto do entretenimento digital, expandindo o escopo de sua validação para além de questões políticas ou sociais [4, 5, 29].

Em comparação com estudos sobre o impacto da crítica de cinema na bilheteria [12, 13], este trabalho se diferenciaria ao focar na cobertura de notícias em um agregador como o Google Notícias e em métricas de popularidade de *streaming* (JustWatch), que refletem o consumo em um ambiente digital e sob demanda, em vez de bilheteria de cinema. A análise de dados de plataformas digitais em tempo real oferece uma perspectiva mais atualizada sobre a dinâmica do consumo de entretenimento.

As implicações dos resultados esperados são significativas. Para a indústria do entretenimento, a demonstração de uma correlação positiva entre notícias e rankings de *streaming* sublinha a importância de estratégias de relações públicas e assessoria de imprensa para gerar visibilidade e impulsionar o consumo. Para os veículos de comunicação, reforça seu papel como formadores de opinião e influenciadores do consumo cultural.

No entanto, as limitações inerentes a um estudo correlacional devem ser reconhecidas. A principal delas é a impossibilidade de inferir causalidade direta. É difícil isolar o efeito da imprensa de outros fatores que contribuem para a popularidade de um título. Além disso, a dependência de dados de *web scraping* pode introduzir desafios relacionados à completude e consistência dos dados, bem como a potenciais vieses algorítmicos do Google Notícias em sua agregação de conteúdo. A generalização

dos resultados para todos os tipos de filmes/séries ou para outras plataformas de notícias/*streaming* deve ser feita com cautela.

5. Reflexões Éticas e Limitações

A coleta de dados através de *web scraping* levanta importantes questões éticas e legais que devem ser cuidadosamente consideradas:

- **Política de Uso e robots.txt:** O estudo reconhece que a maioria dos serviços online não deixa clara sua política de uso para *web scraping*. Embora os arquivos `robots.txt` dos sites utilizados não apresentem restrições explícitas, é fundamental entender que o `robots.txt` é uma diretriz voluntária e não possui força legal [23]. No entanto, ignorá-lo pode ser considerado antiético. A prática recomendada é limitar o número de páginas raspadas durante o desenvolvimento e depuração do código para evitar sobrecarregar os servidores [24].
- **Autorização para Extração Automatizada:** O próprio plano de pesquisa sugere o contato para autorização de extração automatizada. Esta é uma prática ética fundamental: "pergunte educadamente" [24]. Se o projeto requer dados de uma organização específica, o ideal é tentar obter os dados diretamente em formato estruturado, se possível [24]. Para o JustWatch, existe uma API de parceiros que pode ser acessada mediante contrato [16, 17].
- **Direitos Autorais e Uso Justo:** É crucial reconhecer que o *web scraping* pode ser ilegal se os termos e condições do site proibirem especificamente o download e a cópia de seu conteúdo [24]. No entanto, se os dados são publicamente disponíveis (não protegidos por senha) e o objetivo não é roubar conteúdo, copiá-lo palavra por palavra ou monetizá-lo sem permissão, o reuso de conteúdo raspado publicamente é geralmente aceitável para fins de pesquisa pessoal, podendo se enquadrar na provisão de "uso justo" da legislação de propriedade intelectual [24]. O compartilhamento de dados raspados para os quais não se detém os direitos é ilegal [24].
- **Privacidade de Dados:** Embora os rankings e notícias sobre filmes/séries não sejam dados pessoais sensíveis, a atividade do usuário que alimenta os rankings do JustWatch é. É essencial garantir que o estudo utilize apenas dados agregados e anonimizados, sem qualquer possibilidade de identificação individual.
- **Não Quebrar a Internet:** É uma responsabilidade ética garantir que o processo de *scraping* não sobrecarregue os servidores dos sites, causando interrupções de serviço [24]. Ajustar as configurações do *scraper* para permitir um atraso entre as requisições é uma boa prática [24].

O presente estudo, apesar de sua relevância, possui algumas limitações intrínsecas que devem ser reconhecidas:

- **Dependência de Dados de Terceiros:** A pesquisa depende da disponibilidade e da estrutura dos dados fornecidos por JustWatch e Google Notícias. Alterações nas plataformas (e.g., mudanças na estrutura do site, políticas de acesso) podem impactar a coleta de dados.
- **Ausência de Causalidade:** Como um estudo correlacional, não é possível estabelecer uma relação de causa e efeito direta entre a cobertura de notícias e a popularidade. Outros fatores não mensurados podem influenciar ambos.

- **Viés na Coleta de Notícias:** O Google Notícias, como um agregador, possui seus próprios algoritmos de seleção e ranqueamento de notícias, o que pode introduzir vieses na amostra de artigos coletados. Além disso, a cobertura de notícias pode ser influenciada por campanhas de marketing das próprias produtoras, criando uma circularidade.
- **Escopo Limitado:** O estudo se restringe a duas plataformas específicas (JustWatch e Google Notícias) e não abrange a totalidade do ecossistema de *streaming* ou de fontes de notícias (e.g., redes sociais, blogs especializados).
- **Complexidade da Análise de Sentimento:** A análise de sentimento de textos em português pode ser desafiadora e exigir modelos de PLN robustos e adaptados ao domínio do entretenimento, o que pode introduzir imprecisões.

A dependência de dados de plataformas específicas pode introduzir vieses. Por exemplo, os rankings do JustWatch são baseados na atividade de seus usuários, que podem não ser representativos de toda a população de consumidores de *streaming*. Da mesma forma, a seleção de notícias pelo Google Notícias pode não refletir a totalidade da cobertura da imprensa ou pode ter um viés geográfico ou editorial.

A generalização dos resultados para outros mercados, tipos de conteúdo (e.g., documentários vs. filmes de ação) ou plataformas de *streaming* e notícias deve ser feita com cautela. Os padrões de consumo e a influência da mídia podem variar significativamente em diferentes contextos culturais e tecnológicos.

6. Conclusão e Trabalhos Futuros

Este trabalho propôs uma metodologia para investigar a influência da cobertura de notícias nos rankings de popularidade de filmes e séries em plataformas de *streaming*. Ao utilizar dados do JustWatch e do Google Notícias, o estudo visa quantificar a relação entre o "hype" gerado pela imprensa e o consumo de conteúdo audiovisual. A abordagem detalhada para coleta, pré-processamento e análise de correlação, fundamentada em teorias da comunicação como a Agenda-Setting, representa uma contribuição para a compreensão da dinâmica do mercado de *streaming* e do papel da mídia na formação da agenda de consumo.

A pesquisa aqui delineada busca fornecer evidências empíricas sobre como a imprensa, através da cobertura de notícias, pode impactar a visibilidade e a popularidade de filmes e séries. A hipótese de que uma maior cobertura de notícias se correlaciona com uma melhor posição nos rankings de *streaming* é plausível e reforça o papel contínuo da mídia tradicional na era digital. Compreender essa dinâmica é crucial para produtores de conteúdo, plataformas de *streaming* e veículos de comunicação, permitindo-lhes otimizar suas estratégias e entender melhor o comportamento do consumidor.

Com base nos resultados esperados e nas reflexões sobre as limitações, diversas direções para trabalhos futuros são propostas:

- **Análise de Causalidade:** Ir além da correlação, explorando métodos estatísticos mais avançados, como análise de séries temporais (e.g., causalidade de Granger) ou modelos de regressão com variáveis defasadas, para tentar inferir relações causais entre a cobertura de notícias e as mudanças nos rankings.
- **Análise de Sentimento Aprofundada:** Implementar e validar modelos de Processamento de Linguagem Natural (PLN) para uma análise de sentimento

mais granular dos artigos de notícias, investigando como o tom (positivo, negativo, neutro) e a emoção (e.g., entusiasmo, surpresa) da cobertura afetam a popularidade.

- **Expansão das Fontes de Dados:** Incluir outras fontes de notícias (e.g., blogs especializados, redes sociais como Twitter/X, Reddit) e outros rankings de popularidade de *streaming* (e.g., Nielsen Streaming Content Ratings [1, 6]) para uma visão mais abrangente.
- **Fatores Confounding:** Incorporar variáveis de controle na análise, como o orçamento de marketing dos títulos, a presença de grandes estrelas no elenco, ou o impacto de premiações, para isolar melhor a influência da imprensa.
- **Modelagem Preditiva:** Desenvolver modelos de aprendizado de máquina que possam prever a ascensão ou queda de títulos nos rankings de *streaming* com base em métricas de cobertura de notícias.
- **Aspectos Legais e Éticos do Web Scraping:** Aprofundar a pesquisa sobre as implicações legais do *web scraping* em diferentes jurisdições e explorar a viabilidade de estabelecer acordos formais com as plataformas para acesso a dados, garantindo a conformidade e a sustentabilidade da pesquisa.
- **Estudo Qualitativo:** Complementar a análise quantitativa com estudos qualitativos, como entrevistas com profissionais da indústria ou grupos focais com consumidores, para obter *insights* sobre como a imprensa influencia suas decisões de consumo.

Referências

- [1] Nielsen, "Top 10,"[Online]. Available: <https://www.nielsen.com/data-center/top-ten/>. [Accessado em 01/07/2025].
- [2] Kantar IBOPE Media, "Análises e Dados,"[Online]. Available: <https://kantariopemedia.com/conteudo/dados-rankings/>. [Accessado em 01/07/2025].
- [3] M. Castells, *A Sociedade em Rede*. São Paulo: Paz e Terra, 1999.
- [4] M. E. McCombs and D. L. Shaw, "The Agenda-Setting Function of Mass Media,"*Public Opinion Quarterly*, vol. 36, no. 2, pp. 176-187, 1972.
- [5] D. L. Shaw and M. E. McCombs, *The Agenda-Setting Function of Mass Media*. New York: Oxford University Press, 1977.
- [6] TudoCelular, "Streaming fica mais popular do que TV, aponta levantamento da Nielsen nos EUA,"2024. [Online]. Available: <https://www.tudocelular.com/mercado/noticias/n236388-streaming-fica-mais-popular-do-que-tv.html>. [Accessado em 01/07/2025].
- [7] A. Alves, "A influência dos vídeos na construção de identidade dos jovens,"2016.
- [8] G. Gerbner, "Toward a General Theory of Cultural Indicators,"*American Behavioral Scientist*, vol. 19, no. 4, pp. 457-468, 1976.
- [9] Verywell Mind, "What Is Cultivation Theory?,"[Online]. Available: <https://www.verywellmind.com/cultivation-theory-5214376>. [Accessado em 01/07/2025].

- [10] J. G. Blumler and E. Katz, *The Uses of Mass Communications: Current Perspectives on Gratifications Research*. Beverly Hills, CA: Sage, 1974.
- [11] Wikipedia, "Uses and gratifications theory,"[Online]. Available: https://en.wikipedia.org/wiki/Uses_and_gratifications_theory. [Accessado em 01/07/2025].
- [12] Redalyc, "Turismo induzido por filmes: uma revisão sistemática da literatura,"2019. [Online]. Available: <https://www.redalyc.org/journal/4735/473564441011/html/>. [Accessado em 01/07/2025].
- [13] Revista Unitins, "Adaptação cinematográfica e ensino de literatura: uma revisão integrativa,"2020. [Online]. Available: <https://revista.unitins.br/index.php/humanidadeseinovacao/article/view/3302/3242>. [Accessado em 01/07/2025].
- [14] Unisinos, "Binge-watching de séries: quando o guilty pleasure vira padrão de consumo,"2017. [Online]. Available: <https://revistas.unisinos.br/index.php/fronteiras/article/download/22210/60748896/60790687>. [Accessado em 01/07/2025].
- [15] Estudogeral, "Televisão em acção: processo comunicacional mediado como elemento estruturante de mundos coletivos,"2014. [Online]. Available: <https://estudogeral.uc.pt/bitstream/10316/26998/3/Televis%C3%A3o%20em%20a%C3%A7%C3%A3o%20processo%20comunicacional%20mediado%20como%20elemento%20estruturante%20de%20mundos%20coletivos.pdf>. [Accessado em 01/07/2025].
- [16] JustWatch, "JustWatch Streaming API,"[Online]. Available: <https://www.justwatch.com/us/JustWatch-Streaming-API>. [Accessado em 01/07/2025].
- [17] JustWatch, "JustWatch Partner API Documentation,"[Online]. Available: <https://apis.justwatch.com/docs/api/>. [Accessado em 01/07/2025].
- [18] Oxylabs, "Google News Scraper API,"[Online]. Available: <https://oxylabs.io/products/scraper-api/serp/google/news>. [Accessado em 01/07/2025].
- [19] Stanford University, "Detecting Political Bias in News Sources using NLP,"2024. [Online]. Available: https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1224/reports/custom_116661041.pdf. [Accessado em 01/07/2025].
- [20] Media Impact Project, "NLP and the News Media,"[Online]. Available: http://www.mediaimpactproject.org/uploads/5/1/2/7/5127770/nlp-and-media_final.pdf. [Accessado em 01/07/2025].
- [21] QuestionPro, "Correlation Analysis: Definition, Types, Examples, and How to Measure,"[Online]. Available: <https://www.questionpro.com/features/correlation-analysis.html>. [Accessado em 01/07/2025].
- [22] PMC, "Correlation and regression,"2016. [Online]. Available: <https://pmc.ncbi.nlm.nih.gov/articles/PMC5079093/>. [Accessado em 01/07/2025].

- [23] IT Brew, "Why robots.txt is no longer enough to protect against web scraping,"2024. [Online]. Available: <https://www.itbrew.com/stories/2024/03/21/prof-nicholas-vincent-explains-why-robots-txt-is-no-longer-enough-t> [Accessado em 01/07/2025].
- [24] UCSB Library, "Ethics & Legality of Web Scraping,"[Online]. Available: <https://carpentry.library.ucsb.edu/2022-05-12-ucsb-webscraping/06-Ethics-Legality-Webscraping/index.html>. [Accessado em 01/07/2025].
- [25] Aimultiple, "Web Scraping Ethics: A Comprehensive Guide,"[Online]. Available: <https://research.aimultiple.com/web-scraping-ethics/>. [Accessado em 01/07/2025].
- [26] Jornal USP, "Crescimento do streaming modifica o consumo de produções audiovisuais,"2022. [Online]. Available: <https://jornal.usp.br/atualidades/crescimento-do-streaming-modifica-o-consumo-de-producoes-audiovisua> [Accessado em 01/07/2025].
- [27] UFPB, "O jornalismo e as redes sociais: um estudo de caso da cobertura do #ForaMicarla,"2016. [Online]. Available: <http://www.ccta.ufpb.br/ppj/contents/documentos/dissertacoes-2016/dissertacao-cassiana.pdf>. [Accessado em 01/07/2025].
- [28] Gallup, "Following Public Individuals for News Charts,"2023. [Online]. Available: <https://news.gallup.com/poll/506084/following-public-individuals-news-charts.aspx>. [Accessado em 01/07/2025].
- [29] Lumen Learning, "Media Effects Theories,"[Online]. Available: <https://courses.lumenlearning.com/suny-massmedia/chapter/2-2-media-effects-theories/>. [Accessado em 01/07/2025].
- [30] D. L. Shaw and M. E. McCombs, "The Agenda-Setting Function of Mass Media,"*Public Opinion Quarterly*, vol. 36, no. 2, pp. 176-187, 1972.
- [31] Wikipedia, "Cultivation theory,"[Online]. Available: https://en.wikipedia.org/wiki/Cultivation_theory. [Accessado em 01/07/2025].