

Comparação da Cobertura Jornalística Entre Portais de Notícias e YouTube: Um Estudo de Caso do Conflito entre Israel e Palestina

Victor Martins, Jussara M. Almeida, Marcos Gonçalves
Departamento de Ciência da Computação, Universidade Federal de Minas Gerais
Belo Horizonte, Brasil
{victor.martins,jussara,mgoncalv}@dcc.ufmg.br

ABSTRACT

This study systematically investigates how the journalistic coverage of the same event can vary depending on the dissemination platform and the communication vehicle. The research focuses on the coverage of the Israel-Palestine conflict, analyzing content from official news portals and the YouTube channels of major mass-reach media organizations. To this end, we employed Natural Language Processing (NLP) techniques to comparatively characterize both the texts published on the portals and the video transcripts aired on the respective YouTube channels. For sentiment analysis, we utilized the TextBlob and VADER algorithms. Furthermore, we applied text similarity methods, such as the Longest Common Subsequence (LCS) technique and the SBERT (Sentence-BERT) model, to measure the degree of resemblance between headlines and correlated content. This helped to identify when seemingly similar texts actually converge or diverge in their focus. Using BERTopic, it was observed that, despite the existence of similar headlines between portals and videos, the associated texts frequently address distinct themes or emphasize varied aspects of the conflict. This methodological combination allowed us to demonstrate that the journalistic narrative is not uniform and can be significantly influenced by the dissemination platform, even when the base event is the same.

KEYWORDS

platform dissemination, Israel-Palestine conflict, natural language processing, topic modeling, text similarity

1 INTRODUÇÃO

O *YouTube* atua como uma fonte significativa de informações sobre política e assuntos atuais, com vários canais que oferecem diversos estilos de conteúdo e estratégias de engajamento [24]. O vasto conteúdo gerado nessa plataforma oferece uma oportunidade valiosa para investigar como as pessoas percebem e respondem a questões geopolíticas complexas [12]. Contudo, são amplamente conhecidos os efeitos do viés cognitivo dos autores e usuários sobre a produção e recepção de conteúdo [8]. Por conseguinte, esse fenômeno levanta preocupações importantes sobre o direcionamento tendencioso dos sentimentos públicos, disseminação de discurso de ódio, notícias falsas e outras formas de conteúdo prejudicial, alimentadas pelas interações dos usuários. Analogamente, a transmissão de informações enviesadas em sites de notícias pode ocorrer de diversas maneiras, influenciando a percepção pública dos eventos. Como

apontado por Liu [17] e Goudis [10], os jornais podem apresentar viés político-partidário, em que linhas editoriais e visões políticas de um jornalista afetam a cobertura de uma notícia.

Um exemplo de evento de destaque mundial e amplamente noticiado e discutido *on-line* é a guerra entre Israel e Palestina pela Faixa de Gaza. Apesar de historicamente conhecido, esta disputa intensificou-se em 2023, tornando-se desde então tema frequente no noticiário mundial e gerando muita discussão tanto nas mídias tradicionais (i.e., televisão, jornais) quanto em plataformas online. Tomando esta guerra como estudo de caso, esse trabalho visa analisar como a cobertura jornalística varia dependendo da plataforma. Especificamente, analisamos as características textuais de notícias associadas ao conflito e veiculadas tanto nos portais oficiais quanto nos canais do YouTube de importantes veículos de comunicação em massa no mundo. Nosso estudo foca na situação pós-declaração de cessar-fogo, que ocorreu em janeiro de 2025 [4], cobrindo mais de 6 mil notícias divulgadas entre 12 de janeiro e 19 de fevereiro de 2025. Nosso estudo é motivado pela pergunta: *Há diferenças entre a forma como um veículo dissemina informação em diferentes plataformas: no seu portal de notícias e em seu canal no YouTube?*

Para responder a essa questão, empregamos técnicas avançadas de Processamento de Linguagem Natural (PLN) para comparar, de forma sistemática, tanto as notícias publicadas nos portais quanto as transcrições de vídeos veiculados nos canais correspondentes no *YouTube*. Especificamente, realizamos análise de tópicos por meio de análise de sentimentos via *TextBlob* [18] e *VADER* [13] a técnica de última geração *BERTopic* [5], visando identificar os principais temas que emergem em cada plataforma e avaliar o grau de convergência entre os conteúdos publicados por um mesmo veículo.

Nossa abordagem contempla duas perspectivas complementares: (i) uma análise em nível macro, que considera coletivamente todos os conteúdos produzidos por um canal ou portal durante o período de observação; e (ii) uma análise em nível micro, focada em pares de notícias equivalentes — isto é, narrativas que tratam do mesmo acontecimento — publicadas pelo mesmo veículo em ambas as plataformas analisadas.

A ênfase na comparação entre *YouTube* e portais de notícias justifica-se pela diferença substancial nas dinâmicas de produção, curadoria e consumo de informação que caracterizam cada plataforma, as quais influenciam diretamente a forma como os fatos são narrados. Portais de notícias, historicamente, seguem um modelo textual estruturado, ancorado em práticas jornalísticas tradicionais e hierarquias editoriais consolidadas. Já o *YouTube* favorece formatos audiovisuais mais fluidos, visuais e interativos [21], com distribuição fortemente impactada por sistemas de recomendação que priorizam métricas de engajamento [16]. Por outro lado, os

portais tendem a operar em um ecossistema de distribuição mais previsível, guiado pela lógica da busca orgânica e pela seleção editorial [25]. Essas distinções estruturais e algorítmicas podem resultar em diferenças significativas no enquadramento dos eventos e na ênfase atribuída a determinados aspectos do conflito.

Os resultados obtidos indicam variações expressivas na cobertura temática entre as plataformas. A distribuição de notícias de um mesmo veículo não se apresenta de forma homogênea: alguns veículos concentram-se majoritariamente em uma única plataforma, enquanto outros atuam de forma mais equilibrada, embora, mesmo nesses casos, a atenção dispensada a notícias equivalentes difira entre os meios, como sugerem as discrepâncias nas quantidades de conteúdos emparelhados. Observou-se ainda a presença recorrente de narrativas de cunho identitário e pessoal, que desempenham papel central na forma como o conflito é percebido. Tal característica torna-se particularmente relevante em períodos críticos, como o pós-cessar-fogo, quando a estrutura e o tom da comunicação podem influenciar significativamente interpretações acerca da estabilidade do conflito e de suas implicações políticas, sociais e humanitárias.

O restante deste trabalho está organizado como segue: a Seção 2 cobre trabalhos relacionados e suas contribuições. Na Seção 3, descrevemos o processo de coleta de dados e os caracterizamos. A Seção 4 detalha nossa metodologia de análise. Seção 5 apresenta e discute os principais resultados. A Seção 6 menciona algumas das limitações do trabalho. Na Seção 7, apresentamos nossas conclusões.

2 TRABALHOS RELACIONADOS

Diversos estudos já exploraram a análise psicolinguística e de tópicos entre textos de portais ou do *Youtube*. Por exemplo, Liu [17], propõe um *framework* para análise de textos de portais de notícias voltado para tópicos de guerra, com um estudo de caso focado em notícias de 2023 sobre a guerra Israel-Hamas. Os autores empregaram uma análise crítica do discurso em três sites de notícias, e identificaram diferenças no foco e no posicionamento editorial de cada veículo com base em variações lexicais. Diferente desse estudo, nossa pesquisa expande a análise para incluir também transcrições de vídeos do *YouTube*, permitindo uma comparação direta entre o conteúdo textual dos portais e a adaptação desse discurso para plataformas de vídeo. Ainda, Goudis [10] investiga o viés na cobertura jornalística do conflito em 2024, analisando imagens publicadas por 3 veículos distintos. O estudo questionou a ideia de fotografias como "grande verdade", concluindo que os veículos adotam enquadramentos visuais distintos. Nosso estudo se diferencia por focar na estrutura textual dos conteúdos.

Outras abordagens multi-plataforma de modelagem de tópicos já foram propostas. Kloo *et al.* investigam a disseminação de *fake news* e desinformação nazista após a invasão da Ucrânia pela Rússia em 2022 [14]. Eles também se focam na recepção/disseminação do conteúdo por usuários comentaristas através de redes de tópicos. Já em Debnath *et al* se voltam para temas de geoen engenharia e mudanças climáticas [6], enquanto Lee *et al.* compararam textos da *LexisNexis* e da *Web of Science* [11].

Considerando um cenário onde *YouTubers* compartilham e replicam seus vídeos em diferentes plataformas, Li *et al.* estudam como usuários de diferentes plataformas possuem gostos distintos, influenciados pelas formas com que os usuários engajam com o

conteúdo, fazendo com que um conteúdo que performe bem em um domínio não se dê tão bem em outro [15]. Assim como Aldous *et al.*, nosso trabalho busca compreender tópicos e sentimentos associados ao conteúdo de um mesmo conjunto de organizações e às diferentes mídias em que atuam [1].

Apesar de a literatura sobre a cobertura midiática de conflitos internacionais ser vasta, uma lacuna persiste na comparação direta entre a representação de um mesmo evento em ecossistemas de mídia distintos. A maioria dos estudos analisa plataformas de forma isolada, negligenciando a oportunidade de um pareamento direto que revele como as lógicas de produção e consumo de conteúdo influenciam a narrativa. Este trabalho preenche essa lacuna ao realizar um pareamento direto entre textos, examinando como portais de notícias, com sua abordagem editorial tradicional, se comparam a veículos que operam no *YouTube*, uma plataforma caracterizada por uma dinâmica de audiência e engajamento singular. Por meio dessa abordagem, buscamos evidenciar as diferenças e similaridades temáticas e de foco, aprofundando a compreensão sobre a fragmentação do discurso em um cenário de convergência midiática.

3 CONJUNTO DE DADOS

El-Ghawi *et al.* analisam vieses presentes em manchetes relacionadas ao conflito entre Israel e Palestina postadas no *Facebook*. Neste trabalho, eles identificam termos frequentes em posicionamentos anti-pró Palestina/Israel anotados manualmente por uma equipe composta por linguistas especializados em línguas e dialetos árabes [9]. Baseando-se nos resultados obtidos por eles, um conjunto de palavras-chave para guiar a busca por postagens relacionadas ao conflito entre Israel e Palestina. Esse conjunto foi formulado consultas por meio de permutações entre as palavras apresentadas na Tabela 1. Cada consulta segue o formato "[localidade] [complemento]" (por exemplo, "Israel October" ou "Palestine ceasefire"). Além das consultas, foi necessário definir palavras-chave obrigatórias, também apresentadas na Tabela 1, para garantir que os vídeos e notícias coletados estivessem diretamente relacionados ao tema do conflito. Essas palavras-chave funcionam como filtros adicionais, assegurando que os conteúdos encontrados abordem, de forma mais direta, aspectos relevantes do conflito. Como indicado pelos termos de pesquisa, este trabalho se destina a analisar textos em língua inglesa.

Tabela 1: Palavras-chave para coleta de dados

Localidade	Complemento
Israel, Palestine, Hamas, West Bank, israeli, palestinian	ceasefire, Trump, october, massacre, attack, terrorists, jihad, terror, isis, antisemitic, captors, kidnap, zion, settler, occupy, occupation, bombardment, siege, genocide, aggression, displace

A coleta de dados foi realizada em duas etapas principais, buscando garantir a comparabilidade entre as notícias e os vídeos (transcrições) analisados. Primeiramente, as notícias foram coletadas por meio da *API* do *Google News*, utilizando o conjunto de consultas definido acima. Para obter os textos completos das notícias a partir da *URL* fornecida pelo *Google News*, foi utilizada a biblioteca de extração e seleção de artigos, *Newspaper*¹. Partindo

¹<https://newspaper.readthedocs.io/en/latest/>

dessa coleta inicial, foi possível identificar os veículos de comunicação mais frequentemente presentes nos resultados, servindo como base para a busca de vídeos correspondentes nos canais do YouTube.

Para a coleta desses vídeos, adotou-se a abordagem proposta por Dias *et al.*, utilizando a API oficial do YouTube para o procedimento [7]. A fim de restringir os resultados aos veículos de comunicação previamente identificados, a formulação das consultas incluiu explicitamente o nome do veículo na busca, no formato "[localidade] [complemento] [nome do veículo]" aumentando a probabilidade de recuperar vídeos produzidos por esses mesmos veículos (e.g., mesmos jornais). O conteúdo dos vídeos foi então transcrito por meio da *YouTube Transcript API*, seguindo o procedimento descrito por Pereira *et al.* [22].

Ao utilizar as APIs do *YouTube* e *Google News*, os dados aqui acessados são de acesso público. Nenhuma informação privada ou confidencial foi coletada ou analisada.

Os dados foram coletados a partir do dia 20 de março e referiam-se a dados assíncronos, ainda que voltados para um evento recente cujas nuances ainda estavam sendo compreendidas coletivamente durante o período da busca. Esses dados passaram por um processo de limpeza, com a remoção de *stopwords*, a aplicação de lematização e a exclusão de textos com menos de 25 palavras para garantir uma análise mais robusta. Ao final, restaram 1876 transcrições de vídeos do YouTube e 4506 notícias de portais, resultando em um total de 6382 notícias, publicadas por 67 diferentes veículos de comunicação. Muitos desses veículos são jornais de grande circulação no mundo, como ABC, CBS e FOX News, assim como fontes especializadas em Israel e no Oriente Médio, como The Jerusalem Post e Al Jazeera².

A Figura 1 exibe a frequência de veículos por plataforma, mostrando sua distribuição entre portais e *YouTube*. Observa-se que portais possuem uma quantidade significativamente maior de aparições de veículos do que os vídeos, o que sugere uma escolha editorial deliberada de direcionar maior espaço para notícias voltadas para o conflito.

Por outro lado, algumas fontes possuem uma baixa participação de autores. Por exemplo, The Hill (27 aparições em vídeos) e The Jerusalem Post (1 aparição em vídeos) indicam uma menor dependência de múltiplos criadores de conteúdo nessa modalidade. Além disso, LiveNOW from FOX apresenta números extremamente baixos tanto para notícias (9 aparições) quanto para vídeos (73 aparições), o que pode significar uma abordagem mais restrita na autoria do conteúdo publicado.

Comparando-se notícias e vídeos, nota-se que as primeiras são consistentemente mais frequentes. Em algumas fontes, como CNN e Sky News, os valores são relativamente baixos em ambas as categorias, o que pode indicar que esses veículos não se debruçam sobre o tema. Já dentre as fontes principais, podemos destacar que, entre os jornais, os veículos com maior número de aparições são The Jerusalem Post (com 971 apenas no portal, sua principal plataforma de cobertura do conflito), ABC News, CBS News e Al Jazeera. Já no caso do *YouTube*, os números são mais baixos, com destaque para Al Jazeera English, com 250 vídeos.

A diferença entre a dinâmica de produção em vídeos e em notícias textuais é evidente. Os dados indicam que, nos portais, a

cobertura é fortemente concentrada em um número reduzido de veículos com alta atividade editorial, resultando em maior volume de publicações por fonte. Em contraste, no *YouTube* observa-se uma maior dispersão, com um conjunto mais amplo de canais que contribuem de forma menos frequente, refletindo padrões de participação mais fragmentados. Essa distinção sugere diferenças estruturais na forma como o conteúdo é produzido, distribuído e priorizado em cada plataforma, influenciando tanto a visibilidade das narrativas quanto o alcance das informações.

4 METODOLOGIA DE ANÁLISE

Este estudo aplica técnicas de processamento de linguagem natural para analisar as propriedades do conteúdo textual de notícias de portais e de transcrições de vídeos de canais do YouTube associados a importantes veículos de comunicação em massa, com foco em notícias associadas ao conflito entre Israel e Palestina.

Para extração de tópicos, notícias escritas e transcrições de vídeos foram submetidas ao método estado-da-arte de análise de tópicos *BERTopic*. Na combinação de parâmetros escolhida, foi aplicado o modelo de *embeddings all-MiniLM-L6-v2*, UMAP com tamanho de vizinhança 5 e distância mínima de 0.01 entre os pontos. Técnicas próprias do *BERTopic* para redução de ruído e redistribuição de tópicos também foram aplicadas para melhoria dos resultados. Com essas configurações, foram gerados 14 tópicos distintos (melhor descritos na Seção 5.2). A medida de silhueta para o modelo foi de 11,89%. Esta medida indica que, apesar dos esforços para otimizar o modelo, a diferenciação entre os dados pode não ser tão clara assim. Esta classificação sugere que os clusters podem ser um pouco dispersos ou sobrepostos, mas a separação não é totalmente inadequada. Isso faz sentido ao considerar que notícias se tratam de textos longos que, no nosso contexto, se generalizam ao falar sobre um mesmo conflito.

A partir desses métodos, as análises foram feitas em dois níveis.

Em um nível *macro*, todas as notícias publicadas foram consideradas coletivamente. *BERTopic* foi então utilizado para identificar similaridades e diferenças entre os documentos de um mesmo veículo nas duas plataformas. Além disto, focando especificamente nos vídeos compartilhados no YouTube, nós analisamos como as notícias foram recebidas pela audiência, por meio de diferentes métricas de engajamento coletadas da plataforma.

Em um nível *micro*, as mesmas análises foram feitas considerando notícias individuais. Ou seja, nós inicialmente realizamos um processo de pareamento entre textos do portal e transcrições de vídeos, associados ao mesmo veículo, a fim de identificar a mesma notícia (ou seja, o mesmo conteúdo) Para fazer este pareamento, nós identificamos relações de equivalência entre duas notícias publicadas pelo mesmo veículo nas duas plataformas. A relação de equivalência foi definida como verdadeira entre dois textos se ambos tiverem sido veiculados na mesma data e se seus títulos forem semelhantes. Esta definição parte da intuição de que, se as notícias são equivalentes, então cobrem um mesmo evento em uma mesma data e apresentam uma manchete parecida. Para classificar se ambas as manchetes eram semelhantes e questão de texto, foram abordadas duas estratégias.

Na primeira delas, a semelhança entre os títulos foi calculada através da sequência comum mais longa (LCS) [3]. Isso quer dizer que os termos devem aparecer na mesma sequência nas duas *strings*,

²Os dados coletados, assim como as análises realizadas, podem ser requisitados aos autores via e-mail.

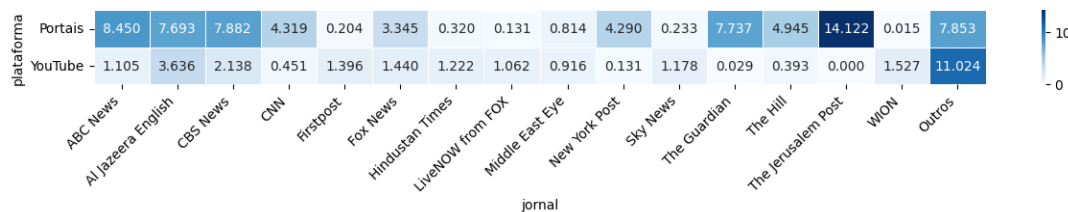


Figura 1: Frequência de Textos por Fonte

mas podem estar intercalados com outros caracteres ou palavras. A taxa de similaridade entre os dois títulos é a razão entre o tamanho da LCS e o tamanho do título mais longo; valores próximos a 0 indicam que ambas as sentenças são completamente diferentes e valores próximos a 1 indicam similaridade máxima.

A segunda estratégia baseou-se em o SBERT (Sentence-BERT) [23], um modelo de *transformer* treinado para produzir representações vetoriais densas (embeddings) de sentenças, permitindo medir sua similaridade sem depender da ordem exata das palavras ou de sobreposições literais. Ao contrário do LCS, que se baseia apenas na correspondência exata de caracteres em sequência, o SBERT capta aspectos semânticos do texto, reconhecendo equivalências mesmo quando as manchetes utilizam sinônimos, variações gramaticais ou estruturas frasais diferentes. Essa capacidade de modelar o significado e o contexto linguístico resulta em uma detecção mais robusta de similaridade textual, especialmente em casos onde a reformulação mantém o mesmo conteúdo informativo.

5 RESULTADOS E DISCUSSÃO

Ao longo dessa Seção, exibiremos os resultados encontrados pelo estudo e exibiremos como estes resultados confirmam ou desacreditam nossas hipóteses, assim como o conhecimento público sobre a guerra entre Israel e Palestina.

5.1 Análise de Sentimentos

A polaridade e a subjetividade das notícias analisadas foram mensuradas por meio dos algoritmos VADER e TextBlob, que quantificam o sentimento de um texto em uma escala de -1 a +1 e o grau de opinião em uma escala de 0 a 1, respectivamente. Através deles, pôde-se notar que a natureza da fonte (YouTube vs Portal) e a perspectiva editorial inerente a cada veículo de comunicação influenciam diretamente a polaridade e a subjetividade de suas reportagens. Ao analisar a polaridade, observou-se que a cobertura do conflito por certas fontes tendia a se concentrar em aspectos específicos, resultando em uma distribuição de sentimentos mais neutra, refletindo um esforço aparente por uma abordagem objetiva. Em contrapartida, outras fontes apresentaram uma polaridade com maior dispersão, indicando que seus textos contêm opiniões mais variadas e polarizadas, que podem ser interpretadas como um reflexo de vieses editoriais ou da inclusão de declarações de diferentes partes envolvidas no conflito.

Adicionalmente, a análise da subjetividade mostrou que todos os veículos analisados possuem uma média de subjetividade próxima a 0, o que sugere uma cobertura predominantemente factual. A seguir, observaremos como a cobertura dessas notícias se distribui ao longo de diferentes tópicos.

5.2 Análise de Tópicos

A Tabela 2 exhibe, para cada um dos tópicos gerados pelo *BERTopic*, os 10 termos mais significativos de cada tópico. As análises a seguir serão realizadas a partir da interpretação desses termos.

Assim, alguns pontos importantes podem ser enfatizados a partir dos resumos gerados. Por exemplo, pode-se observar que, apesar das negociações de cessar-fogo em Gaza, os desafios humanitários permanecem críticos. Hospitais continuam lotados e a reconstrução é um desafio, enquanto milhares de deslocados ainda não têm para onde voltar. No campo político, especialistas discutem se a trégua é temporária ou um possível avanço diplomático. A cobertura midiática reflete disputas narrativas, com acusações de censura e viés, enquanto investigações sobre crimes de guerra ganham força. Apesar da trégua, incidentes violentos pontuais podem ameaçar a estabilidade do acordo. Relatos de massacres e sofrimento civil impulsionam a pressão internacional sobre Israel e Hamas. Enquanto as negociações para troca de prisioneiros continuam sensíveis, com Israel enfrentando resistência interna e o Hamas buscando fortalecer sua posição política. O posicionamento de potências como EUA, Irã e Rússia pode influenciar os desdobramentos, especialmente considerando a escalada militar e a retórica extremista ainda presente. No plano social, famílias separadas pelo conflito enfrentam dificuldades para se reunir, e o trauma da guerra continuará impactando gerações. Apesar da pausa nos combates, a instabilidade e a possibilidade de uma nova escalada são reais. Tais considerações estão de acordo com Alvarez *et al.*, assim como notícias publicadas por portais exteriores aos estudados, que observaram como, mesmo com a declaração inicial de cessar-fogo, a situação permaneceu frágil, com relatos de incidentes isolados de violência e desafios nas negociações para a libertação de reféns e prisioneiros. A crise humanitária na Faixa de Gaza continua sendo uma preocupação significativa, com a comunidade internacional enfatizando a necessidade de garantir a segurança dos civis e fornecer assistência humanitária [2, 19].

5.3 Distribuição dos tópicos por Plataforma

Analisou-se a partir dos tópicos, tornou-se possível identificar quantas notícias foram veiculadas para cada tópico de acordo com cada veículo avaliado, como exhibe a Figura 2. Note que tópicos são referenciados pelos seus nomes, definidos como a concatenação de seu índice e termos mais frequentes.

As visualizações revelam diferenças significativas na cobertura do conflito e, ao analisar veículos específicos, percebe-se que Al Jazeera tem uma alta concentração no YouTube em um único tópico (Tópico 1), relacionado a "Hamas" e "reféns", enquanto sua distribuição nos portais é mais distribuída. Isso pode indicar que o veículo

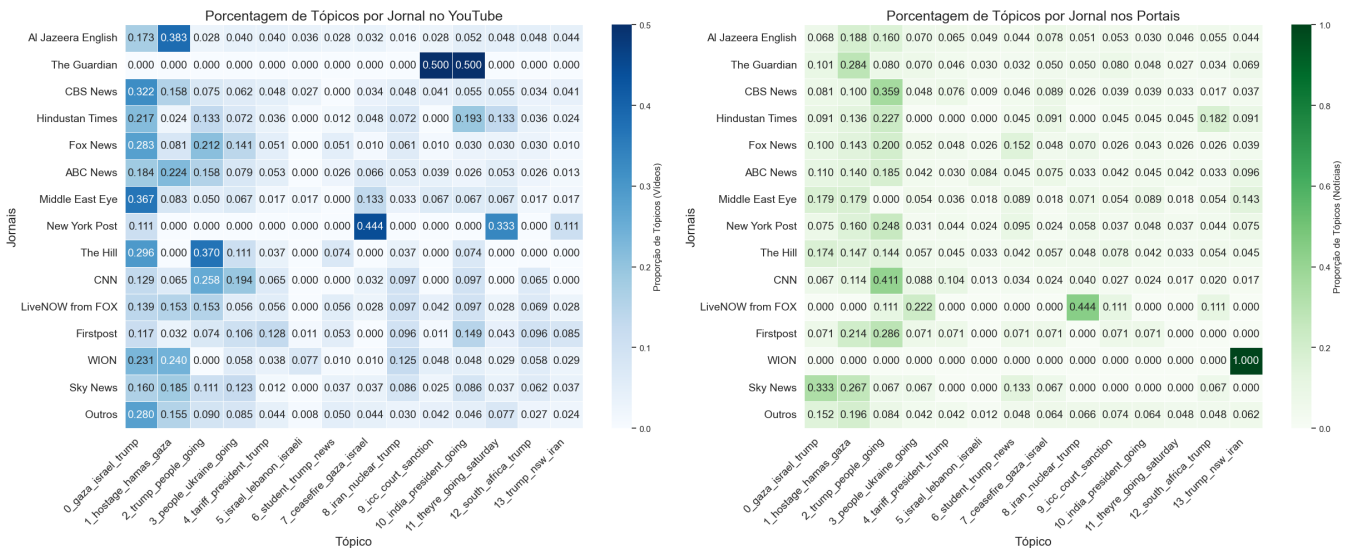


Figura 2: Tópicos por Plataforma

Tabela 2: Resumo dos tópicos do conflito Israel-Palestina

Tópico	15 Palavras Mais Frequentes	# Textos
0	gaza, israel, trump, palestinian, president, people, state, said, hamas, israeli, war, attack, military, peace, conflict	938
1	hostage, hamas, gaza, israeli, palestinian, israel, ceasefire, said, released, deal, negotiation, prisoners, conflict, war, truce	1145
2	trump, people, going, president, fbi, said, dont, thats, know, want, policy, influence, elections, speech, media	899
3	people, ukraine, going, million, want, year, lot, country, dont, president, refugees, crisis, impact, support, help	386
4	tariff, president, trump, going, country, canada, people, india, american, dont, trade, economy, policy, sanctions, business	315
5	israel, lebanon, israeli, hezbollah, lebanese, ceasefire, military, strike, said, attack, border, security, escalation, war, threat	190
6	student, trump, news, new, morning, desk, antisemitism, snow, university, president, protest, campus, debate, incident, media	309
7	ceasefire, gaza, israel, hamas, deal, palestinian, war, israeli, people, peace, negotiations, agreement, future, rebuilding, violence	324
8	iran, nuclear, trump, president, weapon, pressure, iranian, maximum, music, sanction, deal, agreement, war, policy, tensions	369
9	icc, court, sanction, international, israel, trump, said, crime, order, netanyahu, justice, ruling, investigation, trial, charges	329
10	india, president, going, trump, prime, modi, country, minister, thank, want, relations, cooperation, partnership, diplomacy, meeting	331
11	theyre, going, saturday, hostage, music, hell, people, break, gaza, returned, event, emotions, society, cultural, impact	270
12	south, africa, trump, child, compiles, land, people, shad, going, president, policy, rights, reform, political, debate	269
13	trump, nsw, iran, nurse, said, president, pakistan, india, minister, nasa, geopolitics, alliances, international, decisions, leadership	308

utiliza o YouTube para reforçar uma narrativa específica, enquanto nos portais ela adota uma abordagem mais diversificada. Já Fox News e CNN, embora presentes em ambas as plataformas, mostram diferenças na ênfase dos tópicos, refletindo suas linhas editoriais e o possível impacto dos algoritmos do YouTube na amplificação de determinados temas. Outro ponto de destaque é a CBS News, que apresenta uma forte concentração em um único tópico no YouTube, mas uma distribuição mais equilibrada nos portais. No portal de notícias, a WION parece extremamente focada em um único tópico, o que pode indicar uma abordagem especializada sobre um evento específico.

De forma geral, percebe-se que no YouTube há uma tendência de especialização em determinados assuntos, potencialmente buscando atrair nichos específicos. Já nos portais de notícias, a cobertura se espalha por mais tópicos, possivelmente refletindo uma tentativa de oferecer um panorama mais completo da situação.

5.4 Análise de Notícias Equivalentes

A correlação de Pearson[20] entre as medidas de similaridade obtidas pelo LCS e pelo SBERT foi de 0,7569, indicando correlação forte. Para definir um limiar de similaridade, foram selecionados 10 textos aleatórios em cada intervalo de 0,1 (0–0,1; 0,1–0,2; ...; 0,9–1,0). A análise manual realizada pelos autores indicou que o valor de 0,6 na métrica SBERT representa adequadamente a fronteira para considerar dois textos equivalentes. A Tabela 3 apresenta exemplos de equivalência com seus respectivos títulos, valores de similaridade e data e horário de publicação.

Dessa forma, foram identificadas 109 equivalências entre as notícias. A esquerda na Figura 3 exibe como se dá a distribuição delas entre os diferentes veículos. Ao centro, o gráfico exibe quantas dessas equivalências possuem o mesmo tópico associado tanto à notícia no YouTube quanto no portal.

Como podemos ver, os gráficos revelam estratégias diferenciadas para cada um dos veículos. À esquerda, observa-se que CBS News se

Tabela 3: Exemplos de associação de equivalência

Veículo	Título do Vídeo no YouTube	Título da Notícia no Portal	Similaridade	Data
CBS News	Hamas says it will release 6 living hostages, Israel to begin talks on next phase of Gaza ceasefire	Hamas says it will release more hostages than expected this week, including bodies of the Bibas family - CBS News	0.728653	18/02/2025 08:00 – 16:41
Al Jazeera English	Gaza peace negotiations at “critical stage”	Updates: Gaza ceasefire 2nd-phase negotiations begin this week – Israel - Al Jazeera English	0.755729	18/02/2025 06:44
Al Jazeera English	Returning to our homes in Gaza Between Us	Returning to our homes in Gaza - Al Jazeera English	0.767810	18/02/2025 08:00 – 09:30
Fox News	We are looking at the 'last days of Hamas', David Friedman says	David Friedman: We could see the end of Hamas in 'next month or two' - Fox News	0.790887	16/02/2025 08:00 – 22:00
Al Jazeera English	Hamas to handover four Israeli captives' bodies on Thursday	Hamas to release six Israeli captives, hand over four bodies this week - Al Jazeera English	0.796779	18/02/2025 08:00 – 14:33
The Hill	Happening Now: Senators Discuss Trip To Israel From Tel Aviv	Watch live: Senators discuss trip to Israel from Tel Aviv - The Hill	0.905301	17/02/2025 08:00 – 18:52
CBS News	Trump administration holding talks with Russia on Ukraine war, Zelenskyy not attending	Trump administration holding talks with Russia on Ukraine war, Zelenskyy not attending - CBS News	0.942407	17/02/2025 08:00 – 18:13
CBS News	Netanyahu warns "gates of hell" will be opened if Hamas does not release hostages	Netanyahu warns "gates of hell" will be opened if Hamas does not release hostages - CBS News	0.977092	17/02/2025 08:00 – 16:39

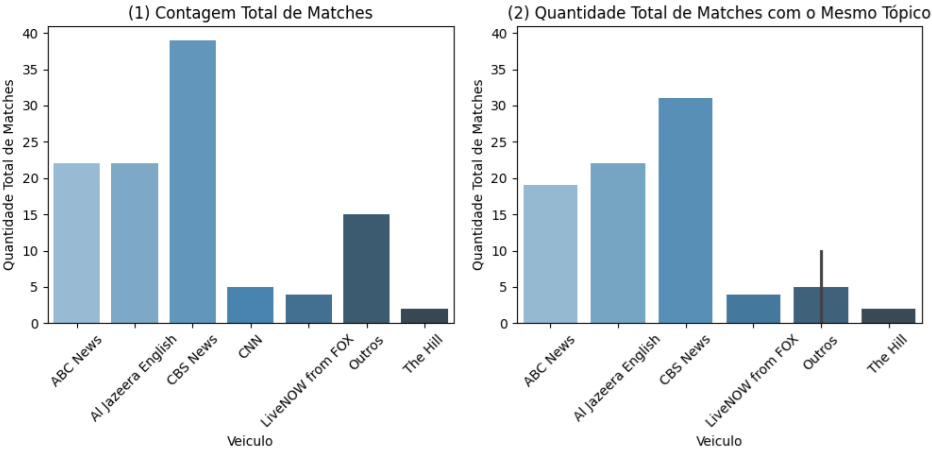


Figura 3: Métricas de Equivalências

destaca por possuir o maior número de notícias pareadas, seguido por Al Jazeera English e ABC News, enquanto outros veículos como LiveNOW from FOX, CNN e The Hill apresentam menor quantidade de equivalências. Essa distribuição sugere que alguns veículos investem mais na replicação de conteúdos entre plataformas, enquanto outros adotam abordagens mais restritas ou diversificadas.

Já à direita observa-se a quantidade de equivalências cujas notícias associadas tratam do mesmo tópico. Al Jazeera e The Hill exibem uma similaridade próxima a 100%, indicando que os conteúdos mantêm alta consistência temática entre os formatos. ABC News e CBS News também demonstram grande alinhamento, ainda que um pouco inferior. Por outro lado, LiveNOW from FOX, CNN e a categoria "Outros" mostram proporções menores de similaridade, sinalizando adaptações que podem variar entre plataformas. Ainda, essa similaridade indica a eficácia do nosso método de pareamento, que, ainda que simples, consegue identificar notícias que tratam de um mesmo evento. Dessa forma, mesmo quando a análise de tópicos indica diferenças entre uma notícia e sua equivalência, isso não necessariamente significa que os conteúdos sejam divergentes em essência. Pelo contrário, a presença de títulos similares aponta

que esses textos e vídeos costumam tratar do mesmo evento central, porém com abordagens distintas. Esse fenômeno revela um comportamento editorial estratégico: um mesmo portal de notícias pode publicar o mesmo conteúdo tanto em seu site escrito quanto em seu canal do YouTube, mas ajustando o enfoque, a profundidade ou o formato para melhor se adequar às especificidades de cada mídia.

Por exemplo, enquanto a notícia no site pode apresentar uma cobertura mais detalhada, analítica e estruturada, o vídeo no YouTube pode explorar elementos visuais, depoimentos ou atualizações mais dinâmicas, gerando variações temáticas que se refletem na análise de tópicos. Essa diferenciação indica que, apesar da replicação do núcleo informativo, os veículos investem em adaptações que potencializam o engajamento e a experiência do público em cada plataforma.

Assim, esse achado evidencia que o método de pareamento utilizado, ainda que simples, é eficaz para capturar essas relações complexas entre conteúdos multimodais. Além disso, destaca a importância de considerar não apenas a similaridade literal entre

títulos e textos, mas também as nuances editoriais que cada canal impõe para diversificar a narrativa, atender às expectativas dos seus públicos e ampliar o alcance da informação. Esse aspecto é particularmente relevante para compreender as estratégias de disseminação da notícia na mídia digital contemporânea, onde a convergência entre formatos coexistem com a necessidade de adaptação temática.

6 LIMITAÇÕES

Apesar da proposta metodológica permitir o pareamento automatizado entre conteúdos de portais de notícia e vídeos do YouTube, algumas limitações importantes devem ser consideradas quanto à validade e robustez dos resultados obtidos. Por exemplo, quanto ao viés de plataforma, ambas as plataformas analisadas apresentam naturezas distintas. Portais jornalísticos costumam seguir linhas editoriais mais estáveis, com apuração e revisão formalizadas, enquanto o YouTube abriga uma diversidade de formatos, tons e autores, incluindo influenciadores, canais alternativos e comentários opinativos. Essa diferença pode introduzir variações significativas na forma como os mesmos eventos são narrados.

Além disso, o pareamento foi realizado com base na similaridade entre títulos e datas de publicação diante da origem comum do conteúdo (mesmo canal). No entanto, essa abordagem pode gerar falsos pares, ou seja, conteúdos que, embora apresentem títulos semelhantes, não tratam exatamente do mesmo evento ou contexto. Além disso, títulos genéricos, ambíguos ou sensacionalistas podem dificultar a distinção semântica necessária para pareamentos precisos. A escolha manual de um limiar para a classificação de equivalência verdadeira foi realizada para mitigar a geração de falsos pares.

A coleta baseada exclusivamente em palavras-chave nos títulos está sujeita a viés de seleção. Palavras neutras podem deixar de captar conteúdos relevantes com carga emocional ou ideológica, enquanto termos mais polarizados podem concentrar os dados em determinadas perspectivas, comprometendo a diversidade informacional do corpus. Nossa seleção de dados buscou balancear ambos os tipos.

Ademais, uma limitação importante observada no contexto do YouTube é a efemeridade das informações publicadas na plataforma, onde conteúdos como vídeos e suas transcrições podem ser removidos, atualizados ou perder relevância com rapidez, dificultando o acesso e a análise contínua ao longo do tempo. Em contraste, os portais noticiosos tendem a manter seus artigos e reportagens disponíveis de forma mais estável e permanente, o que facilita a construção de um arquivo histórico mais consistente para análises futuras. Essa diferença estrutural impõe desafios metodológicos para estudos comparativos, uma vez que a volatilidade do conteúdo no YouTube pode gerar lacunas temporais ou amostras incompletas, afetando a robustez da análise longitudinal da cobertura jornalística.

O uso de BERTopic traz também algumas limitações técnicas. BERTopic depende da qualidade da vetorização dos textos. Ruídos como erros ortográficos, siglas ou termos ambíguos podem comprometer a formação coerente de tópicos.

7 CONCLUSÕES E TRABALHOS FUTUROS

Este estudo apresentou uma metodologia para análise comparativa da cobertura jornalística entre duas plataformas digitais, portais de

notícias e o *YouTube*, aplicada a um estudo de caso sobre o conflito entre Israel e Palestina. Os resultados evidenciaram diferenças substantivas na forma como as notícias são distribuídas e enquadradas em cada ambiente. Identificaram-se variações relevantes na frequência de publicação, na distribuição temática, no tom e na polaridade dos conteúdos. De modo geral, os portais de notícias apresentaram maior volume de cobertura, enquanto o *YouTube* concentrou um número relativamente superior de veículos ativos, direcionados a tópicos mais específicos.

A análise também demonstrou que a cobertura midiática e os debates online abrangem uma ampla gama de questões. O período de cessar-fogo destacou temas como negociações para a libertação de reféns, esforços de reconstrução da região e as relações diplomáticas de Israel e Palestina com países vizinhos. Adicionalmente, tópicos como a influência política de Donald Trump, o aumento do antissemitismo em instituições de ensino superior e a atuação do Tribunal Penal Internacional ilustram como o conflito se insere em uma disputa geopolítica mais ampla, permeada por debates sobre direitos humanos e pela formação da opinião pública global. No momento da escrita deste artigo, o conflito continuou a evoluir, e os tópicos analisados, ainda no começo de 2025, já evidenciavam a complexidade da situação, sugerindo que ela poderia despontar para diversos lados.

A análise de notícias pareadas mostrou que, embora haja conteúdos publicados simultaneamente em ambas as plataformas e associados a tópicos similares, essa correspondência não é absoluta. Isso reflete que, apesar das manchetes semelhantes, os canais ainda optam por focar em tópicos distintos em cada uma das plataformas, adequando seus discursos a elas.

Este estudo apresenta limitações relacionadas à natureza, ao escopo e ao período de coleta dos dados, que contemplaram apenas conteúdos em língua inglesa e concentraram-se majoritariamente em portais de notícias populares em regiões do Norte Global. Pesquisas futuras poderiam incorporar abordagens multilíngues e multirregionais para ampliar a representatividade e a robustez das conclusões. Outros caminhos possíveis rumam a interpretação multimodal dos dados, incluindo não apenas textos, mas também aspectos como imagens, entonação, edição e dinamismo do conteúdo. Apesar dessas restrições, os resultados contribuem para a compreensão das diferenças editoriais entre plataformas digitais, evidenciando como o formato e as estratégias de engajamento influenciam a construção e a percepção das notícias, além de revelar potenciais vieses. Por fim, o procedimento de pareamento entre notícias de diferentes plataformas, embora incipiente, demonstra potencial para aplicações mais amplas em estudos sobre ecossistemas informacionais e circulação de narrativas.

AGRADECIMENTOS

Este trabalho contou com o financiamento de CNPq, FAPEMIG, CAPES e Instituto Nacional de Ciência e Tecnologia em Inteligência Artificial Responsável para Linguística Computacional e Tratamento e Disseminação de Informação (INCT-TILD-IAR).

REFERÊNCIAS

- [1] Kholoud Khalil Aldous, Jisun An, and Bernard J. Jansen. 2023. What really matters?: characterising and predicting user engagement of news postings using

- multiple platforms, sentiments and topics. *Behaviour and Information Technology* 42, 5 (2023), 545–568.
- [2] Artur Alvarez, Luisa Belchior, and Wesley Bischoff. 2025. *Israel e Hamas chegam a acordo por cessar-fogo na guerra em Gaza, anuncia Catar*. <https://g1.globo.com/Acesso em 13 fev. 2025>.
- [3] Alberto Apostolico and Concettina Guerra. 1987. The longest common subsequence problem revisited. *Algorithmica* 2 (1987), 315–336.
- [4] Raffi Berg. 2024. What we know about the Gaza ceasefire deal. *BBC News* (2024). <https://www.bbc.com/news/articles/cy5klgy5zv0o> Acesso em 02 de fevereiro de 2025.
- [5] Muriël de Groot, Mohammad Aliannejadi, and Marcel R. Haas. 2022. Experiments on Generalizability of BERTopic on Multi-Domain Short Text. arXiv:2212.08459
- [6] Ramit Debnath, Pengyu Zhang, Tianzhu Qin, R. Michael Alvarez, and Shaun D. Fitzgerald. 2024. Deciphering public attention to geoengineering and climate issues using machine learning and dynamic analysis. arXiv:2405.07010
- [7] Aline Dias, Richardy Tanure, Jussara Almeida, Helen Lima, and Carlos Ferreira. 2024. Análise da Percepção do Uso de Cigarros Eletrônicos no Brasil por meio de Comentários no YouTube. In *Proceedings of the 30th Brazilian Symposium on Multimedia and the Web*. 45–53.
- [8] Nimrod Dvir, Elaine Friedman, Suraj Commuri, Fan Yang, and Jennifer Romano. 2023. Words That Stick: Predicting Decision Making and Synonym Engagement Using Cognitive Biases and Computational Linguistics.
- [9] Yousra El-Ghawi, Abeer Marzouk, and Aya Khamis. 2024. LexiconLadies at FIG-NEWS 2024 Shared Task: Identifying Keywords for Bias Annotation Guidelines of Facebook News Headlines on the Israel-Palestine 2023 War. In *Proceedings of The Second Arabic Natural Language Processing Conference*. 561–566.
- [10] Nikolaos Alexandros Goudis. 2024. *Beyond the lens: An analysis of CNN, Al Jazeera and protothema.gr news photographs of the 2023 Israel-Palestine war*. Univerzita Karlova, Fakulta sociálních věd, Praga, Tchèqueia.
- [11] Kyeo Re Lee Jang Hyun Kim Haein Lee, Seon Hong Lee. 2023. ESG Discourse Analysis Through BERTopic: Comparing News Articles and Academic Papers. *Computers, Materials & Continua* 75, 3 (2023), 6023–6037.
- [12] Simon Hofmann, Christoph Sommermann, Mathias Kraus, Patrick Zschech, and Julian Rosenberger. 2024. Hate Speech and Sentiment of YouTube Video Comments From Public and Private Sources Covering the Israel-Palestine Conflict. *19th International Conference on Wirtschaftsinformatik* (2024).
- [13] C.J. Hutto and Eric Gilbert. 2014. VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 8. AAAI, 216–225.
- [14] Ian Kloof, Iain J. Cruickshank, and Kathleen M. Carley. 2024. A Cross-Platform Topic Analysis of the Nazi Narrative on Twitter and Telegram during the 2022 Russian Invasion of Ukraine. *Proceedings of the International AAAI Conference on Web and Social Media* 18, 1 (May 2024), 839–850.
- [15] Xinhang Li, Zhaopeng Qiu, Jiacheng Jiang, Yong Zhang, Chunxiao Xing, and Xian Wu. 2023. Conditional Cross-Platform User Engagement Prediction. *ACM Trans. Inf. Syst.* 42, 1, Article 6 (Aug. 2023), 28 pages.
- [16] Alexander Liu, Siqi Wu, and Paul Resnick. 2024. How to Train Your YouTube Recommender to Avoid Unwanted Videos.
- [17] Yankai Liu. 2024. A Corpus-Based Critical Discourse Analysis of News Reports on the 2023 Israel-Hamas War. *Journal of Linguistics and Communication Studies* 3, 3 (Aug. 2024), 70–84.
- [18] Steven Loria. 2014. *TextBlob: Simplified Text Processing*. <https://textblob.readthedocs.io/> Versão usada: 0.15.3. Acesso em 15 de fevereiro de 2025..
- [19] The Observer. 2024. *The Observer view: Shaky ceasefire is no victory for Netanyahu amid suffering of Gaza and hostages*. https://www.theguardian.com/commentisfree/2024/nov/30/the-observer-view-shaky-ceasefire-is-no-victory-for-netanyahu-amid-suffering-of-gaza-and-hostages?utm_source=chatgpt.com Acesso em 13 fev. 2025.
- [20] Karl Pearson. 1895. Notes on regression and inheritance in the case of two parents. *Proceedings of the Royal Society of London* 58 (1895), 240–242.
- [21] Limor Peer and Thomas B. Ksiazek and. 2011. YouTube and the challenge to Journalism. *Journalism Studies* 12, 1 (2011), 45–63.
- [22] Iury Pereira, Juan Avelar, Cláudio Silva, and Daniel Barbosa. 2024. Análise de Sentimentos em Vídeos do Youtube sobre Polarização Política: Uma Abordagem Híbrida Baseada em Reconhecimento de Entidades. In *Anais do XIII Brazilian Workshop on Social Network Analysis and Mining*. 221–227.
- [23] Iryna Reimers, Nils e Gurevych. 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Association for Computational Linguistics, Hong Kong, China, 3982–3992.
- [24] Maud Reveilhac. 2024. YouTube as an information source on politics and current affairs: Supply- and demand-side perspectives. *First Monday* 29 (07 2024).
- [25] David Tewksbury and Jason Rittenberg. 2012. *News on the Internet: Information and Citizenship in the 21st Century*.