

14º Seminário Nacional Sociologia & Política

Saberes em Confluência

GT 15 - Desafios teóricos, metodológicos e éticos em pesquisas com/de mídias e plataformas digitais

BERTopic para estudos de enquadramento no campo da Comunicação Política: uma análise comparativa entre Revista Oeste e Carta Capital no X no ano de 2025

Vitoria Prá¹

INTRODUÇÃO

A consolidação dos espaços informacionais digitais e a ascensão das novas mídias impuseram ao campo da Comunicação Política desafios teóricos e metodológicos, na medida em que os processos comunicacionais se tornaram mais complexos e fragmentados. Nielsen (2014) já apontava a necessidade do campo se reorganizar para ser capaz de estudar as transformações nas lógicas institucionais e a forma como os atores constroem sentido nesses novos ambientes informacionais. Isso sugere uma abertura para teorias e metodologias mais diversas, além de agendas de pesquisa que não se limitam a mudanças imediatas no comportamento político dos indivíduos. Ou seja, as constantes transformações digitais não impactam apenas comportamentos imediatos, mas também alteram as lógicas pelas quais as informações são produzidas, processadas e difundidas. O que influencia comportamentos coletivos e produz novos sentidos sobre as questões que envolvem a experiência humana.

¹ Mestranda do Programa de Pós Graduação em Comunicação da Universidade Federal do Paraná, na linha de Comunicação e Política. Email: vitoriapra@ufpr.br

Compreender como essas novas lógicas informacionais se relacionam com os processos políticos exige que o campo da Comunicação Política mantenha seu repertório teórico e metodológico em constante atualização. Parte disso passa pela necessidade de ampliar métodos e técnicas voltados à análise de vastos corpora digitais. Ao se propor a estudar as interações entre atores políticos, instituições e meios de comunicação, estudiosos frequentemente se deparam com a necessidade de processar volumes de dados que ultrapassam a capacidade humana. Diante desse cenário, o campo da Comunicação Política tem incorporado progressivamente métodos computacionais para o processamento de grandes volumes de dados digitais (Theocharis e Jungherr, 2020).

Entre as técnicas mais amplamente adotadas para a análise de textos nas pesquisas brasileiras em ciências humanas e sociais destaca-se a Classificação Hierárquica Descendente (CHD), também conhecida como método de Reinert. Desenvolvida originalmente por Max Reinert (1990). A CHD parte de uma perspectiva lexicométrica, segundo a qual o vocabulário utilizado por um enunciador expressa os mundos lexicais que estruturam seu discurso. Em termos operacionais, mede a intensidade de presença dos vocábulos em diferentes partes do corpus, identificando associações entre termos por proximidade e intensidade (Cervi, 2018). No Brasil, a difusão da CHD foi fortemente impulsionada pelo software Iramuteq (Camargo e Justo, 2013) que, gratuito e sem exigência de conhecimento de linguagem de programação, facilitou o acesso a métodos automatizados de análise lexical e contribuiu para sua rápida incorporação em diferentes áreas do conhecimento.

É importante ressaltar que a CHD é, em sua natureza, uma técnica de clusterização voltada à identificação de mundos lexicais, verificando como os léxicos se organizam e se associam a outros, permitindo a análise do contexto da presença dos termos (Cervi, 2018). Sua aplicação busca revelar os universos de sentido que atravessam o discurso de um ou mais enunciadores e, portanto, é mais adequadamente aplicada quando o corpus de análise é de um único universo temático ou proveniente de um cenário específico; e o objetivo é identificar diferentes construções discursivas que existem nesse universo. Quando a finalidade da pesquisa é outra, como identificar e categorizar tematicamente grandes volumes de documentos de assuntos diversos, as técnicas de modelagem de tópicos podem ser mais adequadas. Nesse campo, o Latent Dirichlet Allocation (LDA) é o modelo mais amplamente utilizado: trata-se de um modelo probabilístico que pressupõe que cada documento é composto por uma mistura de tópicos latentes, os quais são definidos por distribuições de probabilidade sobre o vocabulário do corpus (Blei, Ng e Jordan, 2003). Em outras palavras,

enquanto a CHD é adequada para mapear os mundos lexicais e estruturas discursivas de um enunciador ou conjunto de enunciadores em um único universo temático ou contextual, o LDA é uma técnica capaz de identificar os diferentes temas presentes em um grande volume de documentos. Embora operem sob lógicas distintas, ambas as abordagens baseiam-se em representações do tipo bag-of-words, tratando os documentos como coleções não ordenadas de palavras nas quais apenas a frequência e a coocorrência dos termos importam. Essa escolha de representação apresenta desempenho satisfatório em duas situações: quando os documentos são longos o suficiente para gerar padrões lexicais estáveis; e quando o corpus versa sobre um conjunto relativamente coerente de temas, de modo que as mesmas palavras-chave tendam a se repetir com frequência ao longo dos textos.

O problema surge quando o corpus não oferece as condições que essas representações pressupõem. As transformações digitais reconfiguraram as formas pelas quais a informação é produzida, distribuída e consumida (Chadwick, 2013). No campo da Comunicação Política, isso se traduz em uma disputa crescente por visibilidade e atenção nas plataformas digitais. Com a migração dos fluxos informativos para as redes sociais, os veículos passam a depender delas para atrair tráfego (Ju et al., 2014; Nielsen e Ganter, 2022), competindo com outros veículos, influenciadores e lideranças políticas num ambiente em que o alcance é condicionado por estruturas algorítmicas. Nesse contexto, as chamadas publicadas nas redes sociais deixam de ser apenas resumos do conteúdo jornalístico e passam a ser elaboradas em função de sua capacidade de gerar engajamento e atrair tráfego para os sites. Kumpel, Karnowski e Keyling (2015), em revisão sistemática da literatura sobre compartilhamento de notícias nas redes sociais, identificaram que fatores textuais e de conteúdo das publicações dos veículos são determinantes centrais de sua propagação, e que o tráfego oriundo das redes sociais se tornou indispensável para que os veículos alcancem seus públicos. Nesse cenário, estudar as chamadas que os veículos publicam nas redes sociais importa para compreender como eles disputam atenção e constroem sentido sobre a realidade social.

Se as manchetes são, portanto, importantes objetos de análise para estudos de enquadramento em plataformas digitais, o corpus resultante é composto por um grande volume de textos muito curtos e sobre assuntos altamente variados entre si. É nesse cenário que as técnicas baseadas em bag-of-words encontram suas maiores limitações: com pouquíssimas palavras por documento e alta diversidade temática, a informação de frequência e coocorrência disponível por documento é escassa, e diferentes formulações textuais que expressam conteúdos semanticamente próximos dificilmente são agrupadas de forma coerente

(Qiang et al., 2020). Diante desse cenário, outras técnicas derivadas do LDA foram propostas, com o objetivo de resolver o problema da limitação de dados esparsos em textos curtos, mas que continuam com a premissa de bag-of-words (Costa e Duarte, 2019). O presente trabalho, no entanto, propõe a utilização do BERTopic (Grootendorst, 2022) como ferramenta auxiliar na análise de enquadramento de chamadas de notícias publicadas nas redes sociais, uma vez que, diferentemente das abordagens baseadas em frequência e coocorrência de palavras, analisa cada documento como uma unidade semântica, representando-o a partir do significado contextual do conjunto de suas palavras.

As seções seguintes desenvolvem esse argumento em duas etapas. Primeiro, apresenta-se o BERTopic, descrevendo seu funcionamento e discutindo suas potencialidades e limitações para os estudos da Comunicação. Em seguida, propõe-se uma aplicação empírica da técnica por meio de uma análise comparativa das publicações da Revista Oeste e da CartaCapital no X ao longo de 2025. Tratando-se de veículos com orientações ideológicas em polos opostos do espectro político, mas com estratégias digitais semelhantes — os tweets de ambos consistem, em geral, nas chamadas das notícias acompanhadas do link de acesso ao conteúdo completo —, o objetivo é apresentar o BERTopic como uma ferramenta que facilite a organização de corpora extensos e ofereça um mapa para análises de enquadramento.

O BERTOPIC COMO FERRAMENTA PARA ESTUDOS DE COMUNICAÇÃO POLÍTICA

Grootendorst (2022) propõe o BERTopic como uma resposta ao problema de representações do estilo bag-of-words: de não levar em consideração as relações semânticas entre as palavras. Ou seja, não consideram que, por vezes, a mesma palavra pode representar significados distintos quando aplicada em diferentes construções lexicais ou que frases distintas podem fazer parte do mesmo contexto mesmo sem repetir palavras. Por exemplo: as frases “Ministro da Fazenda aumenta imposto sobre as bets” e “Fernando Haddad anuncia maior carga tributária sobre as casas de apostas” têm o mesmo significado apesar da falta de compartilhamento lexical entre si. Quando o corpus contém apenas frases curtas como essas, sem o apoio de documentos mais amplos — como notícias completas, nas quais palavras relevantes tenderiam a se repetir com maior frequência —, os modelos de bag-of-words apresentam dificuldade em identificar a similaridade semântica entre elas.

Assim, o BERTopic é apresentado como uma ferramenta de modelagem de tópicos que gera representações vetoriais densas de cada documento completo (cada manchete, tweet ou frase é um documento, por exemplo), os chamados embeddings, a partir de modelos de linguagem previamente treinados que capturam o significado do texto em sua totalidade, pela combinação dos significados das palavras que o compõem. Segundo Grootendorst (2022, p. 1, tradução nossa), “as propriedades semânticas dessas representações vetoriais permitem que o significado dos textos seja codificado de tal forma que textos similares se encontram próximos no espaço vetorial”. Isso significa que, uma vez que "imposto" está semanticamente próximo de "tributária" e "bets" está próximo de "apostas" nesse espaço, o modelo produzirá representações vetoriais próximas para as duas frases do exemplo.

A partir dessas representações vetoriais, o BERTopic opera em três etapas sequenciais. Na primeira, os vetores — que possuem centenas de dimensões — são comprimidos em um espaço de menor dimensionalidade, viabilizando o processamento computacional sem perder as características do espaço multidimensional. Na segunda etapa, os documentos comprimidos são agrupados por densidade: o algoritmo identifica os agrupamentos a partir da proximidade dos vetores no espaço, reunindo em um mesmo tópico os documentos semanticamente similares. Documentos que não apresentam proximidade suficiente com nenhum agrupamento são isolados como ruído e alocados em um tópico residual, convencionalmente denominado tópico -1. Essa característica representa, simultaneamente, um diferencial e uma limitação do BERTopic. Por um lado, ao não forçar todos os documentos a pertencerem a algum tópico, o modelo se diferencia de outras ferramentas de clusterização que forçam todos os documentos a pertencerem a algum tópico, “limpando” ruídos que poderiam dificultar a coesão e a compreensão. Por outro lado, documentos alocados no tópico -1 podem conter conteúdos relevantes para determinados objetos e objetivos de pesquisa, o que torna sua análise uma etapa recomendável em muitos contextos.

A terceira etapa serve, a princípio, para facilitar a interpretação dos tópicos. O modelo reúne todos os documentos de cada tópico em um único documento, e mede a importância de cada palavra para o significado geral do tópico e, com isso, identifica quais as palavras que mais distinguem aquele tópico dos demais (Grootendorst, 2022). Para além de permitir a identificação rápida do tema geral de cada tópico, essa etapa fornece um mapa analítico para estudos de enquadramento, uma vez que apresenta quais termos são mais relevantes para a formação de um tópico. No caso de análises comparativas, como o caso do exemplo deste trabalho, permite identificar os diferentes temas distinguíveis em cada tópico entre autores

diversos. Ou seja: quando dois ou mais autores ou grupos de autores falam da mesma coisa, quais os léxicos que cada um mobiliza que se diferenciam dos demais? Essa etapa, no entanto, não se resolve por si só com análises de enquadramento, e sim oferece direcionamentos para onde o pesquisador deve olhar, facilitando o estudo de corpora extensos.

Assim, o BERTopic apresenta potencial relevante para estudos de enquadramento no campo da Comunicação Política, especialmente em dois aspectos. O primeiro diz respeito à escala analítica: ao automatizar a identificação de agrupamentos temáticos em grandes volumes de dados, a ferramenta permite ampliar o escopo das análises, seja pela extensão do recorte temporal, seja pela ampliação de enunciadores, tornando possíveis análises comparativas mais abrangentes e longitudinais. O segundo aspecto diz respeito à sensibilidade semântica da ferramenta. Matthes e Kohring (2008) já sublinhavam o risco de viés do pesquisador na identificação de enquadramentos e apontavam o potencial dos métodos computacionais para reduzir a interferência analítica na identificação de enquadramentos. Os autores reconheciam, contudo, que as ferramentas disponíveis à época — baseadas em bag-of-words — apresentavam limitações para capturar relações semânticas mais complexas, identificando padrões de coocorrência lexical sem necessariamente apreender a proximidade de sentido entre textos de vocabulário distinto. O BERTopic avança nesse ponto ao operar sobre representações semânticas densas, sendo capaz de agrupar documentos lexicalmente distintos, mas tematicamente próximos.

Há, no entanto, limitações importantes que precisam ser consideradas. A primeira, e mais fundamental, é de ordem conceitual: tópicos não são enquadramentos. Na definição de Entman (1993), um enquadramento implica seleção e saliência de determinados aspectos de uma realidade percebida, envolvendo dimensões que vão além da identificação do tema abordado. Ali e Hassan (2022), em levantamento de estudos de análise de enquadramento assistida por computador, identificaram que a grande parte das abordagens computacionais trata tópicos como proxies de frames, o que representa uma inconsistência conceitual relevante, uma vez que saber sobre o que um texto fala não equivale a saber como enquadra aquilo de que fala. Essa distinção não é resolvida pelo BERTopic, apenas facilitada. Isto é, a ferramenta é adequada para identificar agendas temáticas e saliências, mas a análise de enquadramento propriamente dita requer esforços qualitativos complementares sobre os tópicos identificados.

Uma segunda limitação, apontada pelo próprio Grootendorst (2022), é de ordem estrutural: o BERTopic pressupõe que cada documento pertence a um único tópico, o que é

uma simplificação adequada para unidades textuais curtas — como manchetes e tweets —, mas pode representar uma limitação quando aplicado a documentos mais extensos, que frequentemente articulam múltiplos temas simultaneamente. Isso reforça a adequação da ferramenta ao tipo de corpus do presente trabalho, mas também limita seu alcance em outros contextos analíticos. Além disso, como utiliza um modelo transformer pré-treinado, essa restrição afeta a replicabilidade exata da pesquisa, na medida em que o modelo é atualizado e aprimorado. Por fim, a utilização do BERTopic pressupõe ao menos familiaridade básica com a linguagem de programação Python, bem como conhecimento dos parâmetros da ferramenta, uma vez que os resultados são sensíveis às configurações escolhidas, como o número mínimo de documentos por tópico e os parâmetros dos algoritmos de redução de dimensionalidade e clusterização. Ainda que sua operacionalização não demande formação avançada em ciência da computação e há tutoriais detalhados na internet, inclusive no site de Grootendorst no github², essa exigência pode representar uma barreira de entrada real para pesquisadores das humanidades sem experiência prévia com programação.

Ao levar em consideração as potencialidades e limitações discutidas até aqui, esse trabalho propõe a aplicação do BERTopic numa perspectiva comparada de dois enunciadores posicionados em polos opostos do espectro ideológico.

APLICAÇÃO EMPÍRICA: ANÁLISE COMPARATIVA ENTRE REVISTA OESTE E CARTA CAPITAL NO ANO DE 2025

Como exercício de aplicação empírica, realizamos uma análise comparativa entre a Revista Oeste e a Carta Capital, a partir da coleta do X. Os veículos foram escolhidos por compartilhar um modelo de conteúdo similar — as publicações na rede se resumem às manchetes das notícias publicadas no site, acompanhadas dos links para acesso ao texto completo — e, ao mesmo tempo, terem alinhamentos editoriais opostos, posicionando-se como mídias alternativas ideologicamente orientadas: a Revista Oeste identificada à direita do espectro, e a Carta Capital, à esquerda.

O conceito de mídia alternativa adotado neste trabalho se alinha à perspectiva de Holt, Figenschou e Frischlich (2019) sobre a necessidade de atualizá-lo, entendendo que a perspectiva anterior limitava as possibilidades de estudos empíricos sobre este fenômeno nos

² Disponível em <https://maartengr.github.io/BERTopic/index.html>. Acesso em 28 de mar 2026

contextos contemporâneos. Para os autores, as mídias alternativas devem ser compreendidas em contraste às mídias tradicionais e por uma perspectiva multilevel, que pode incluir a identificação de produtores alternativos, conteúdos alternativos e/ou processos alternativos de produção e distribuição, mas não necessariamente associados a movimentos sociais de esquerda, como inicialmente eram referenciadas (Fiorucci, 2011). Desse modo, torna-se um conceito guarda-chuva, não ideológico, a ser mobilizado menos como uma ferramenta de diagnóstico e mais como um mapa teórico para investigações empíricas sobre como e em que nível o posicionamento alternativo se reflete nesses processos midiáticos e suas consequências para o debate público (Holt, Figenschou e Frischlich, 2019).

O processo metodológico aplicado se valeu de algumas etapas. Primeiro, foram coletados os tweets dos perfis de cada um dos portais, utilizando como filtro as 5000 publicações com maiores taxas de engajamento³ no ano de 2025. Esse procedimento foi adotado para equalizar o volume de publicações analisados de cada portal, uma vez que a Revista Oeste publicou 2,5 vezes mais do que a Carta Capital no período, evitando super representação no corpus. O próximo passo consistiu na junção do corpus, com o identificador do emissor, e limpeza dos textos (retirando caracteres especiais e hiperlinks). Em seguida, o modelo foi treinado na ferramenta Google Colab, utilizando linguagem Python, limitando-se ao parâmetro padrão nas etapas de geração dos embeddings e clusterização, aplicando-se a presença mínima de 50 documentos como critério para a formação de um tópico. Essa decisão viabiliza a análise proposta neste trabalho, por um lado, mas por outro significa a exclusão de temáticas mais dispersas e locais.

Na etapa de identificação dos termos mais representativos de cada tópico, foram utilizados parâmetros que retiram stopwords e diminuem a presença de palavras muito frequentes ou muitos termos similares entre si. O código completo e o banco de dados utilizados neste trabalho estão disponíveis em repositório público no GitHub.⁴

A modelagem foi realizada para o corpus completo, sem considerar os diferentes autores, com o objetivo de encontrar tópicos em comum e tornar possível comparar a presença de cada tema em cada portal. No entanto, para identificar possíveis diferenças de enquadramento, a terceira etapa do BERTopic, que consiste na identificação das palavras mais representativas de cada tópico, foi recalculada considerando as classes (os veículos). A modelagem resultou em 32 tópicos distintos, além do tópico -1, que reuniu os documentos

³ Engajamento é calculado pela razão entre as interações e o volume de seguidores

⁴ Disponível em : https://github.com/vitoriapra/paper_bertopic_14sns

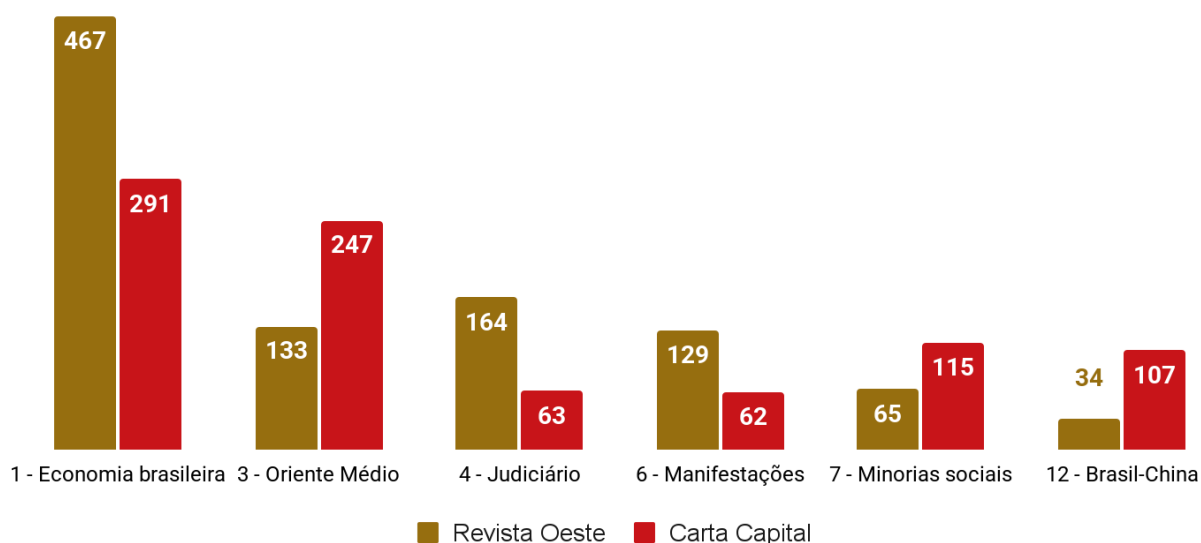
que não foram encaixados em nenhum tópico. O grande volume de documentos no tópico -1 (5036 documentos, sendo 2414 da Revista Oeste e 2622 da Carta Capital) pode ser explicado pelos parâmetros adotados, principalmente a definição mínima de 50 documentos para formar um tópico. Com isso, buscamos observar apenas os temas que mais apareceram no ano de 2025 nos dois portais. Análises mais aprofundadas e específicas poderiam reduzir essa limitação e/ou analisar os outliers alocados para o tópico -1. Para fins deste trabalho, seguiremos com os 32 tópicos identificados pelo modelo, que podem ser observados no Anexo 1.

O primeiro resultado mais imediato é a diferença entre os portais no volume de documentos de alguns tópicos (gráfico 1). O tópico 1, que reúne manchetes relacionadas à economia brasileira, com destaque para a questão da fraude do INSS, é o que conta com maior volume de documentos no corpus total, sendo o tópico mais comum da Revista Oeste e o segundo da Carta Capital. Isso sugere, assim, uma disputa de narrativas em torno do tema. A frequência 60% superior do tópico na Revista Oeste revela uma atenção maior da direita ao caso, uma vez que o governo em exercício tende a ser mais responsabilizado por escândalos de corrupção, ainda que as evidências indiquem que a fraude foi iniciada no governo anterior.

O tópico 3, identificado a partir dos termos representativos como se tratando de conflitos do Oriente Médio, aparece com uma frequência quase duas vezes superior (+85%) na Carta Capital, enquanto o tópico 4, que reúne documentos associados ao judiciário, aparece com uma frequência 160% maior na Revista Oeste. Além de indicar a maior atenção de cada um dos portais aos temas, também sugere maior engajamento da audiência. Em outras palavras, é possível que o filtro das 5000 publicações com maior engajamento tenha excluído mais publicações da Revista Oeste que estariam no tópico do Oriente Médio, enquanto há um volume muito superior da Carta Capital que passou pelo filtro. O tópico 6 trata de manifestações e protestos, e aparece 2x mais na Revista Oeste que na Carta Capital. Um olhar rápido nos tweets classificados nesse tópico revela que a grande maioria diz respeito às manifestações bolsonaristas pela anistia dos presos do 8 de janeiro, o que explica a maior atenção do portal da direita ao tema. Por outro lado, os tópicos 7 e 12 são mais comuns na Carta Capital: enquanto o tópico 7, relacionado a questões de minorias sociais, alinha-se à orientação editorial do veículo⁵, o tópico 12 revela a grande atenção dada à relação Brasil-China, uma das principais bandeiras do governo Lula.

⁵ “Visão: Manter CartaCapital como a maior referência em jornalismo progressista no Brasil, em qualquer plataforma.” Trecho retirado da página ‘Princípios’ no portal da Carta Capital. Acesso em 29 de março de 2026.

Gráfico 1 - Tópicos com a maior diferença de frequência entre os dois portais



Fonte: Autora (2026)

Além das diferenças de saliência temática, os resultados também indicam caminhos para identificação de enquadramentos divergentes, por meio da análise das palavras mais representativas de cada tópico dos dois veículos (quadro 1). A presença do termo “corrupção” entre as palavras representativas da Revista Oeste no tópico 1, por exemplo, sugere um enquadramento que busca dar evidência a suspeitas de corrupção do governo Lula, sabidamente um tema no qual os governos petistas tem desvantagem perante a opinião pública. A menção à “Epstein” nos termos representativos da Carta Capital no tópico 5 aponta para possíveis associações entre as lideranças políticas americanas, em especial Trump, e os “Epstein Files”. Ao buscar na base completa do corpus, encontramos manchetes da Carta Capital como “Democratas publicam novas fotos de Epstein com Trump” e “‘Não é minha assinatura’, diz Trump sobre suposta carta a Jeffrey Epstein”.

Quadro 1 - Tópicos em que as palavras representativas indicam saliências divergentes entre os veículos

Tópico	Palavras representativas Oeste	Palavras representativas Capital
1 - Economia Brasileira	fraude, fraudes, bilhões, impostos , corrupção	bilhões, fraude, fraudes, crédito, banco
5 - Política americana	trump, convite, mail, presidente, washington	trump, musk, epstein , americanos, aliado

6 - Manifestações	manifestantes, manifestação, apoiadores, paulista, pacificação	manifestantes, milhares, protestam, protestos, apoiadores
7 - Minorias sociais	trans, mulheres, sexual, racismo, lgbt	trans, racismo, transfobia , estupro , diversidade
10 - Regulação das redes sociais	sociais, regulação, censura , internet, google	sociais, regulação, google, internet, responsabilização
17 - Educação	escolas, educação, linguagem , professora, professores	escolas, professores, educação, ensino, estadual
28 - Vacinas	vacina, obrigatória , bebês , crianças , cartão	vacina, febre, laboratório, autismo, vacinada

Fonte: Autora (2026)

O tópico 6 aponta para a presença do termo “pacificação” em notícias associadas às manifestações a favor da anistia, indicando um tipo de enquadramento dos eventos que salienta o argumento de que a anistia seria o único caminho para a pacificação do Brasil polarizado. A leitura de chamadas de notícias com esse termo revela a utilização da estratégia de jornalismo declaratório, que transforma declarações de figuras públicas em notícias e viabiliza espaços para declarações desinformativas, ataques a adversários e promoção de interesses pessoais (Pollacchinni e Benetti, 2025), como é mostrado no Quadro 2::

Quadro 2 - Exemplos de tweets que contém o termo “pacificação”

Veículo	Texto do tweet
Revista Oeste	Tarcísio defende anistia como saída para 'pacificação'
Revista Oeste	Temer diz que Moraes vai cumprir 'papel de pacificação' na redução de penas no 8/1
Revista Oeste	Aldo Rebelo defende anistia: 'Única forma de pacificação do Brasil'
Revista Oeste	Temer defende atuação de Moraes no âmbito do 8 de janeiro: quer a 'pacificação do país'
Revista Oeste	'Num sopro de energia, deputados começam a se proteger dos abusos do Supremo e tentam avançar com a anistia, o único trilho para a pacificação.' Leia a reportagem de Silvio Navarro na Revista Oeste

Fonte: Autora (2026)

Outro tópico em que a diferença de enquadramento fica evidente pela leitura das palavras representativas é o tópico 7, que reúne as publicações de manchetes sobre questões de minorias sociais. Além de “diversidade” ser um termo representativo apenas para a Carta

Capital, a presença dos termos “estupro” e “transfobia” aponta para notícias que denunciam crimes contra minorias sociais. Já a presença do termo “racismo”, na lista de ambos os veículos, exige uma segunda etapa de análise que identifique como a questão aparece em cada um deles (Quadro 3)..

Quadro 3 - Exemplos de tweets que contém o termo ‘racismo’, por veículo⁶

Veículo	Texto do tweet
Revista Oeste	Influenciador nega racismo e sai em defesa de Bolsonaro
Revista Oeste	Ministra das Mulheres de Lula é alvo de denúncia por assédio e racismo
Revista Oeste	Deputado aciona PGR contra Lula por suposto crime de racismo
Revista Oeste	Em visita a quilombo, Barroso diz que o Judiciário luta contra o 'racismo estrutural'
Carta Capital	A articulação de deputada bolsonarista para mudar a Lei do Racismo
Carta Capital	Cármem Lúcia denuncia racismo sofrido por ministra do TSE
Carta Capital	Coluna O Brasil não disfarça mais o seu racismo
Carta Capital	PUC-SP suspende por 30 dias alunos denunciados por racismo contra estudantes cotistas
Carta Capital	Justiça do Trabalho condena farmácia por racismo contra funcionária
Carta Capital	Entidade pede à Justiça indenização de R\$ 750 milhões da CBF e da Conmebol por casos de racismo
Carta Capital	Pentágono manda apreender em bibliotecas militares material sobre diversidade e antirracismo

Fonte: Autora (2026)

A seleção dos casos noticiados por cada veículo revelam os sentidos que são construídos em torno da questão. No caso da Revista Oeste, a questão do racismo aparece como instrumento de disputa política: os únicos casos de racismo denunciados são aqueles supostamente cometidos por membros do governo Lula. Os outros dois casos indicam relativização da questão, seja pela negação de um caso (a acusação não aparece, apenas a negação), seja pelo uso de aspas no termo ‘racismo estrutural’, descredibilizando o conceito. Ao analisar matérias da revista A Folha de São Paulo, Vianna (2011) identifica o uso das aspas nesse contexto como forma de destacar um termo e se distanciar dele ao mesmo tempo, passando ao leitor a função de compreender porquê o termo está destacado e elaborar sua própria interpretação sobre ele. Assim, o emprego das aspas esvazia o conceito,

⁶ Foram identificados 4 tweets do corpus total em que a Revista Oeste mobiliza o termo “racismo”, todos apresentados na tabela 3. No caso da Carta Capital, o termo apareceu em 20 tweets do corpus. Pela limitação deste trabalho, são apresentados aqui alguns exemplos ilustrativos. Outros tweets podem ser conferidos no banco de dados completo disponibilizado no repositório.

descredibilizando a tradição teórica que compreende o racismo estrutural conceito importante para a compreensão da realidade social.

A denúncia do racismo como parte de disputa política também aparece na Carta Capital, na chamada “A articulação de deputada bolsonarista para mudar a Lei do Racismo”, que intencionalmente qualifica a deputada em questão como bolsonarista. Além disso, exemplos como “Coluna | O Brasil não disfarça mais o seu racismo” e “Pentágono manda apreender em bibliotecas militares material sobre diversidade e antirracismo” apontam para uma atuação de jornalismo crítico e denunciativo. Esse exercício revela como a identificação de termos representativos em um agrupamento temático não esgota a análise, mas oferece direcionamentos de como aprofundá-la.

O tópico 10, que reúne as chamadas que tratam da regulação das redes sociais, apresenta indicativos de julgamentos morais opostos, nos termos de Entman (1993): a regulação associada à censura, por parte da Revista Oeste, e à ideia de responsabilização, por parte da Carta Capital. Por fim, os tópicos “Educação” e “Vacinas” sugerem enquadramentos da Revista Oeste que se alinham a retóricas da direita bolsonarista. A presença do termo “linguagem” aponta para a saliência do debate sobre linguagem neutra, enquanto termos como “obrigatória”, “bebês” e “crianças” sugere enquadramentos que reforçam a insegurança em relação à vacinação, sobretudo pela operacionalização de pânico envolvendo bebês e crianças.

CONSIDERAÇÕES FINAIS

Esse trabalho apresentou o BERTopic como ferramenta auxiliar para estudos de enquadramento no campo da Comunicação Política, discutindo seu funcionamento, potencialidades e limitações. A análise da aplicação empírica demonstrou que a modelagem de tópicos permite identificar saliências temáticas e oferece direcionamentos para a análise de enquadramento, ao revelar léxicos que indicam marcadores ideológicos, sendo particularmente útil em estudos sobre enunciadores ou temáticas polarizadas. Os exemplos discutidos ilustraram como a leitura das palavras representativas de cada tópico por veículo constitui um ponto de partida analítico relevante para identificação de enquadramentos.

Algumas limitações do exercício empírico merecem ser explicitadas. O filtro das 5000 publicações com maior engajamento por portal, adotado para equalizar os volumes do corpus, pode ter enviesado a representação temática de cada veículo, favorecendo conteúdos

com maior apelo junto a cada audiência. O parâmetro mínimo de 50 documentos para a formação de um tópico, por sua vez, excluiu temáticas mais dispersas ou locais, o que explica em parte o volume expressivo de documentos alocados no tópico residual (tópico -1). Para análises futuras, percebe-se a necessidade de explorar esse material ou ajustar os parâmetros de clusterização para ampliar a cobertura temática. Por fim, uma limitação inerente ao uso de modelos de linguagem pré-treinados diz respeito à replicabilidade longitudinal da análise: como o BERTopic depende de modelos de embeddings que são periodicamente atualizados e aprimorados, a aplicação do mesmo procedimento em momentos distintos pode gerar agrupamentos diferentes para um corpus idêntico, na medida em que as representações semânticas produzidas pelos modelos evoluem com o tempo. Essas limitações, no entanto, não invalidam os resultados, uma vez que análises não computadorizadas, que utilizam codificadores humanos, também apresentam impossibilidade de replicação perfeita.

Por fim, defende-se a utilização do BERTopic como parte de um desenho metodológico mais amplo, que combine métodos computacionais, quantitativos e qualitativos. Nessa direção, Guo et al. (2022) propõem um framework de cinco etapas para a análise computacional de enquadramento que situa a modelagem de tópicos na primeira etapa, como ferramenta de exploração indutiva do corpus, orientando as decisões analíticas subsequentes. Ou seja, a contribuição do BERTopic está em organizar grandes volumes de dados e mapear os agrupamentos temáticos predominantes, fornecendo ao pesquisador um panorama estruturado a partir do qual é possível selecionar tópicos prioritários, formular hipóteses sobre enquadramentos e conduzir análises qualitativas direcionadas. A identificação dos frames propriamente ditos requer novas etapas metodológicas, podendo incluir tanto a definição dos enquadramentos a partir da leitura manual dos documentos que contenham saliências fornecidas pelo BERTopic, quanto pela aplicação de outras técnicas computacionais.

REFERÊNCIAS

ALI, Mohammad; HASSAN, Naeemul. A survey of computational framing analysis approaches. In: CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING (EMNLP), 2022, Abu Dhabi, Emirados Árabes Unidos. **Proceedings** [...]. Stroudsburg: Association for Computational Linguistics, 2022. p. 9335-9348. Disponível em: <https://aclanthology.org/2022.emnlp-main.633>. Acesso em: 21 mar. 2026.

BLEI, David M.; NG, Andrew Y.; JORDAN, Michael I. Latent Dirichlet Allocation. **Journal of Machine Learning Research**, v. 3, p. 993–1022, jan. 2003. Disponível em: <https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf>. Acesso em: 21 mar. 2026.

CAMARGO, Brígido Vizeu; JUSTO, Ana Maria. IRAMUTEQ: um software gratuito para análise de dados textuais. **Temas em Psicologia**, Ribeirão Preto, v. 21, n. 2, p. 513-518, dez. 2013. DOI: <https://doi.org/10.9788/TP2013.2-16>. Disponível em: http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1413-389X2013000200016&lng=pt&nrm=iso. Acesso em: 21 mar. 2026.

CERVI, Emerson Urizzi. Análise de conteúdo automatizada para conversações em redes sociais online: uma proposta metodológica. In: ENCONTRO ANUAL DA ANPOCS, 42., 2018, Caxambu. **Anais [...]**. [S.L.], 2018.

CHADWICK, Andrew. **The hybrid media system**: politics and power. Oxford: Oxford University Press, 2013.

COSTA, Matheus; DUARTE, Julio. Modelagem de tópicos para textos curtos: uma revisão. In: ESCOLA REGIONAL DE BANCO DE DADOS (ERBD), 15, 2019, [S.L.]. **Anais [...]**. Porto Alegre: SBC, 2019. Disponível em: <https://sol.sbc.org.br/index.php/erbd/article/view/8478/8379>. Acesso em: 21 mar. 2026.

ENTMAN, Robert M. Framing: toward clarification of a fractured paradigm. **Journal of Communication**, Oxford, v. 43, n. 4, p. 51-58, dez. 1993. DOI: <https://doi.org/10.1111/j.1460-2466.1993.tb01304.x>.

FIORUCCI, Rodolfo. A nova geração do jornalismo crítico: mídia alternativa. **Diálogos – Revista do Departamento de História e do Programa de Pós-Graduação em História**, Maringá, v. 15, n. 2, p. 455-481, maio/ago. 2011.

GROOTENDORST, Maarten. BERTopic: neural topic modeling with a class-based TF-IDF procedure. **arXiv preprint**, 11 mar. 2022. DOI: 10.48550/arXiv.2203.05794. Disponível em: <https://arxiv.org/abs/2203.05794>. Acesso em: 11 fev. 2026.

GUO, Lei et al. Proposing an open-sourced tool for computational framing analysis of multilingual data. **Digital Journalism**, [s.l.], v. 10, n. 9, p. 1467–1485, 2022. DOI: <https://doi.org/10.1080/21670811.2022.2031241>.

HOLT, Kristoffer; FIGENSCHOU, Tine Ustad; FRISCHLICH, Lena. Key dimensions of alternative news media. **Digital Journalism**, v. 7, n. 7, p. 860–869, 2019. DOI: <https://doi.org/10.1080/21670811.2019.1625581>.

IAROVICI, Edith; AMEL, Ruthanne. The strategy of the headline. **Semiotica**, v. 77, n. 4, p. 441–459, 1989.

JEONG, Jae-Hwa; JEONG, S. Ho; CHYI, Hsiang Iris. Will social media save newspapers? **Journalism Practice**, v. 8, n. 1, p. 1-17, 2014. DOI: <https://doi.org/10.1080/17512786.2013.794022>.

KÜMPEL, Anna Sophie; KARNOWSKI, Veronika; KEYLING, Till. News sharing in social media: a review of current research on news sharing users, content, and networks. **Social Media + Society**, [S.L.], v. 1, n. 2, jul./dez. 2015. DOI: <https://doi.org/10.1177/2056305115610141>.

MAHMOUD, Ashraf F. A. et al. A comparative evaluation of LDA, NMF, and BERTopic: analyzing perplexity and coherence metrics. **Ingénierie des Systèmes d'Information**, v. 30, n. 12, p. 3163–3169, 2025. DOI: 10.18280/isi.301208. Disponível em: <https://www.iieta.org/journals/isi/paper/10.18280/isi.301208>. Acesso em: 11 fev. 2026.

MATTHES, Jörg; KOHRING, Matthias. The content analysis of media frames: toward improving reliability and validity. **Journal of Communication**, v. 58, n. 2, p. 258–279, jun. 2008.

NIELSEN, Rasmus Kleis. Political communication research: new media, new challenges, and new opportunities. **MedieKultur: Journal of Media and Communication Research**, v. 30, n. 56, p. 5-22, 2014.

NIELSEN, Rasmus Kleis; GANTER, Sarah Anne. **The power of platforms: shaping media and society**. Oxford: Oxford University Press, 2022.

POLLACCHINNI, Rafaela; BENETTI, Marcia. Jornalismo declaratório e tipologia das declarações: asserções avaliativas, descritivas e prescritivas. **Mídia e Cotidiano**, Niterói, v. 19, n. 2, p. 237-262, maio/ago. 2025. Disponível em: <https://periodicos.uff.br/midiaecotidiano/article/view/66473>. Acesso em: 17 abr. 2026.

QIANG, Jipeng et al. Short text topic modeling techniques, applications, and performance: a survey. **arXiv**, 13 abr. 2019. Disponível em: <https://arxiv.org/abs/1904.07695>. Acesso em: 11 fev. 2026.

REINERT, Max. ALCESTE, une méthodologie d'analyse des données textuelles et une application: Aurélia de G. de Nerval. **Bulletin de Méthodologie Sociologique**, n. 28, p. 24-54, 1990.

THEOCHARIS, Yannis; JUNGHERR, Andreas. Computational social science and the study of political communication. **Political Communication**, v. 38, n. 1-2, p. 1–22, 2021. DOI: 10.1080/10584609.2020.1833121. Disponível em: <https://www.tandfonline.com/doi/full/10.1080/10584609.2020.1833121>. Acesso em: 11 fev. 2026.

VIANNA, Rodolfo. **Jornalismo, ironia e “informação”**. 2011. 220 f. Dissertação (Mestrado em Linguística Aplicada e Estudos da Linguagem) – Pontifícia Universidade Católica de São Paulo, São Paulo, 2011.

ANEXO 1 - Tópicos identificados pelo BERTopic

Tópico	Freq Oeste	Palavras representativas Oeste	Freq Capital	Palavras representativas Capital
1	467	fraude, fraudes, bilhões, impostos, corrupção	291	bilhões, fraude, fraudes, crédito, banco
2	342	portugal, artigo, mundo, sanções, pior	314	indonésia, recorde, portugal, seleção, economia
3	133	israel, netanyahu, nuclear, elimina, israelense	247	israel, irã, israelenses, netanyahu, israelense
4	164	jurista, judiciário, advogada, tribunais, juízes	63	advogado, advogados, justiça, juízes, racistas
5	116	trump, convite, mail, presidente, washington	87	trump, musk, epstein, americanos, aliado
6	129	manifestantes, manifestação, apoiadores, paulista, pacificação	62	manifestantes, milhares, protestam, protestos, apoiadores
7	65	trans, mulheres, sexual, racismo, lgbt	115	trans, racismo, transfobia, estupro, diversidade
8	65	santos, evangélicos, vaticano, bispo, igreja	114	vaticano, igreja, jesus, evangélicos, católica
9	97	perde, g7, lula, corina, machado	73	g7, viaja, exigirá, disputas, encontro
10	88	sociais, regulação, censura, internet, google	62	sociais, regulação, google, internet, responsabilização
11	94	paulista, jorjy, sarah, paula, jorjão	54	amaral, eduardo, paradeiro, jorjy, cassar
12	34	china, chinesa, jinpíng, xi, importar	107	china, tarifas, xi, exportações, jinpíng
13	66	cirurgia, hospital, médico, médicos, médica	68	hospital, cirurgia, câncer, médico, médicos
14	71	venezuela, paraguai, salvador, colômbia, cartel	51	venezuela, salvador, colômbia, deportados, latina
15	51	prisão, condenado, condenada, vota, condenar	65	prisão, condenado, 17, zambelli, condena
16	45	2026, candidatura, presidência, eleições, eleição	68	2026, eleição, reeleição, candidato, eleições
17	26	escolas, educação, linguagem, professora, professores	75	escolas, professores, educação, ensino, estadual
18	32	senado, senador, senadores, auditável, caravanas	57	senado, senador, votar, senadores, votações
19	39	julgamento, acusação, acusados, defesa, financeiro	42	julgamento, turma, boate, penal, condenação
20	65	filipe, defesa, entrevistar, embargo, acusa	11	filipe, martins, entrevistar, instrumentos, vinhos

21	51	jornalistas, jornalista, imprensa, repercute, jornalismo	22	jornalistas, jornalista, imprensa, jornalismo, ameaçou
22	54	ministros, ministro, jantar, convoca, acreditam	17	ministros, brumadinho, hungria, insistência, oficiala
23	27	congresso, fechar, presidentes, congressistas, transnacional	44	congresso, guardião, congressistas, afronta, inimigos
24	37	multa, pagar, milionária, honorários, desembargador	32	multa, pagar, indenização, supremacista, mil
25	5	confirmam, ucrânia, envenenamento, ucraniana, guerra	62	ucrânia, russos, istambul, russa, drones
26	45	vídeo, vídeos, áudio, netflix, divulga	18	vídeos, vídeo, netflix, youtuber, censurar
27	21	argentina, argentinos, pobreza, argentino, índice	37	argentina, greve, argentinos, fmi, argentino
28	25	vacina, obrigatória, bebês, crianças, cartão	33	vacina, febre, laboratório, autismo, vacinada
29	34	policiais, polícia, pcc, policial, criminosos	23	policiais, policial, polícia, gritzbach, rio
30	32	flux, julgamento, incompetência, voto, opõe	24	flux, julgamento, diverge, voto, incompetência
31	35	pgr, manifesta, queixa, recomenda, parcialidade	20	pgr, manifesta, queixa, financiadores, kakay
32	31	toffoli, jato, master, vorcaro, alberto	20	toffoli, lava, corpus, habeas, vasectomia