Computational Humanities Research

www.cambridge.org/chr

A language modeling approach to identifying Russian information confrontation in Colombia

Joe Parson[1,2] and Adriana Jaramillo[3,4]

[1]Florida International University, USA; [2]Elder Research Inc., USA; [3]Elliott School of International Affairs, The George Washington University, USA and [4]Independent Scholar, USA

Research Article

Cite this article: Parson Joe and Adriana Jaramillo. 2026. "A language modeling approach to identifying Russian information confrontation in Colombia." *Computational Humanities Research*, **2**:e12, https://doi.org/10.1017/chr.2026.10031

Received: 30 January 2025
Revised: 30 March 2026
Accepted: 28 April 2026

Keywords:
Colombian media; computational text analysis; narrative mimicry; Russian information operations; sentiment analysis; topic modeling

Corresponding author:
Joe Parson;
Email: Joe@Vasseur.io

Abstract

This study integrates multilingual text analytics with time-series modeling to examine how geopolitical narratives emerging around the 2022 Russia–Ukraine conflict align across Russian and Colombian media, drawing on a corpus spanning 2013–2023 that also encompasses the 2014 conflict period. Drawing on a corpus of more than 38,000 Spanish- and Russian-language news articles (2013–2023), we employ a six-stage pipeline combining multilingual NER, document-level sentiment scoring, `e5-large-instruct` embeddings, BERTopic clustering, and a semi-automated human-LLM labeling workflow. Narrative salience and tone were aggregated to weekly, log-normalized series and analyzed using linear regression and Granger causality tests (lags 1–4 weeks), providing a descriptive view of temporal coupling rather than evidence of direct causal influence.

The period surrounding the 2022 invasion produced the clearest and most coherent patterns. Across four macro-narratives (*Security & Conflict*, *Diplomacy*, *Economy*, and *Politics & Society*), Russian and Colombian coverage exhibited sequential alignment in which shifts in narrative volume were often followed by shifts in evaluative tone. Thirty-two of forty-four target–topic pairs displayed significant lag structures prior to false-discovery correction ($p < 0.05$), with 25 surviving Benjamini–Hochberg adjustment. Categories with lower geopolitical salience (*Technology*, *Health*, and *Off-topic*) displayed less consistent temporal coupling, though this contrast is offered as descriptive context rather than a formal control. While these results do not imply coordinated messaging, they indicate recurring forms of *narrative mimicry*, understood here as patterned convergence in framing and sentiment across media ecosystems.

By demonstrating how transformer-based semantic representations can be integrated with human-in-the-loop interpretation and classical time-series analysis, this study contributes a reproducible workflow for tracing narrative trajectories at scale. The approach provides a methodological foundation for examining how geopolitical frames circulate, mutate, and align across linguistic and regional boundaries, offering new possibilities for computational humanities research on transnational discourse.

Plain language summary

What we did?

We collected more than 38,000 news articles and official statements published in Spanish and Russian between 2013 and 2023. Using modern language-model tools, we grouped the articles into themes, measured whether each story sounded positive or negative, and tracked how these themes rose or fell over time. We then used statistical tests (including a method called "Granger causality" which identifies whether changes in one time series tend to come before changes in another) to see whether Russian news patterns were followed by similar changes in Colombian reporting.

What we found?

During the period surrounding the 2022 Russia–Ukraine invasion, we observed a consistent pattern in four major topic areas: security, diplomacy, politics, and the economy. In these themes, increases in Russian coverage were often followed one week later by increases in Colombian coverage on the same topic. In many cases, Colombian stories also adopted a more negative tone about two weeks after Russian coverage had shifted in that direction. This pattern was less consistent in areas, such as health or technology, though we note this comparison is descriptive rather than a formal test of uniqueness. These findings do not show direct influence, but they do indicate a repeated short-lag alignment between the two media systems.

Why it matters?

The results suggest that narratives prominent in Russian state-backed media sometimes align temporally with Colombian coverage in geopolitically salient domains. Rather than

This article was awarded Open Materials badge for transparent practices. See the Data availability statement for details.

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0), which permits unrestricted re- use, distribution and reproduction, provided the original article is properly cited.

demonstrating direct influence, the study provides a reproducible framework for identifying moments of cross-ecosystem narrative convergence and for examining how geopolitical frames are reframed, echoed, or contested across linguistic and regional boundaries.

Introduction

This study investigates how geopolitical narratives articulated in Russian state-backed media are echoed, reframed, or temporally aligned within Colombia's domestic information environment. Drawing on a multilingual corpus of over 38,000 articles published between 2013 and 2023, the research integrates natural language processing (NLP), semantic clustering, and temporal analysis to examine patterns of narrative alignment between RT en Español, Sputnik Mundo, and Colombian news outlets. Rather than claiming direct causality or coordinated information transfer, the analysis focuses on discursive convergence and narrative mimicry. That is, instances in which Colombian media exhibit framing patterns that resemble, echo, or temporally follow those observed in Russian coverage.

The methodological framework identifies candidate narrative alignments and then evaluates their temporal sequencing using linear regression and Granger causality analysis. Importantly, Granger causality is used here only to test predictive temporal precedence, not to infer intentional influence. As such, all findings are interpreted as suggestive, not confirmatory. This distinction is central to the study's epistemological stance: the goal is not to prove influence operations, but to illuminate how narratives circulate, mutate, and resonate across media ecosystems.

While grounded in computational techniques, this study contributes to computational humanities by using large-scale textual analysis to identify patterns in narrative formation, affective tone, and discursive alignment that would be difficult to trace through close reading alone (Jockers 2013). In this methodological tradition, computational models are valuable not because they replace humanistic interpretation, but because they help surface large-scale regularities that can then be examined in their historical, cultural, and rhetorical context – a position Piper (2018) develops through the concept of enumerative reading and Underwood (2019) through the complementarity of distant and close analysis in historical scholarship. Grimmer, Roberts, and Stewart (2022) make an analogous argument in the social science context, emphasizing that computational text methods require continuous grounding in domain theory and human interpretive judgment; Baden et al. (2022) further identify the interpretation gap as a central methodological challenge in political text analysis, arguing that pattern detection must be anchored in substantive theoretical frameworks (Baden et al. 2022; Grimmer, Roberts, and Stewart 2022; Piper 2018; Underwood 2019). Our use of clustering, sentiment analysis, and temporal modeling therefore serves an interpretive rather than reductive function: these tools help identify candidate sites of convergence that require subsequent contextual reading and qualitative adjudication. The central concept of *narrative mimicry* is thus understood not as simple lexical overlap, but as a pattern of shared narrative trajectories shaped by ideological resonance, institutional distrust, and the symbolic structures of geopolitical discourse. Framed in this way, the question of how state-backed narratives circulate and acquire discursive traction across media ecologies is not only computational, but fundamentally interpretive.

This question is particularly salient in Latin America, where traditions of anti-imperialist rhetoric (Gobat 2013), institutional fragility (Levitsky and Murillo 2013), and polarized media systems (Lupu, Ramírez Bustamante, and Zechmeister 2020) create fertile ground for foreign narratives to be taken up or repurposed. Colombia, long a strategic U.S. ally and a regional hub for security cooperation, has emerged as a focal point for Russian Spanish-language media expansion (Brandt and Wirtschafter 2022; Rouvinski 2022). Prior work documents how Russian outlets frame international affairs through narratives that challenge Western institutions, emphasize multipolarity, and appeal to audiences skeptical of U.S. foreign policy (Pomerantsev 2014; Yablokov 2015). These dynamics make Colombia a compelling site for examining how and under what conditions narrative frames prominent in Russian media become discursively aligned with local coverage, not as proof of imported propaganda, but as potential components of broader domestic meaning-making processes.

To study these processes, this article introduces a hybrid computational–interpretive pipeline combining multilingual named entity recognition (NER), document-level sentiment alignment, topic clustering, and hierarchical narrative construction. A key methodological innovation is the use of a fine-tuned Nous Hermes 2 model (Llama 2 7B variant) deployed on an M1 Ultra GPU for semi-automated topic labeling. These generative large language model-generated labels were iteratively reviewed, debated, and refined through human-in-the-loop (HITL) coding, following best practices in qualitative pattern coding and analytical triangulation (Saldaña 2021). This interaction between machine inference and human judgment reflects a core methodological principle in computational humanities: large-scale pattern detection is most productive when paired with meaning-centered analysis grounded in historical and cultural context.

The findings reveal recurring patterns of temporal and thematic convergence between Russian and Colombian media, especially within macro-narratives related to diplomacy, security and conflict, regional politics, and the economy. These alignments are interpreted as patterns consistent with processes, such as reframing, shared media routines, or ideological consonance, rather than as evidence of direct coordination. By situating these patterns within broader regional dynamics and discursive traditions, the study advances our understanding of how geopolitical narratives travel, acquire local inflections, and embed themselves within distant media ecosystems.

Geopolitical context

Colombia occupies a pivotal position within Latin America due to its geographic location, longstanding political alignment with the United States, and central role in regional security cooperation. With coastlines on both the Pacific and Caribbean, proximity to the Panama Canal, and a history of collaboration with U.S. counternarcotics and counterinsurgency initiatives, Colombia has become a hub within hemispheric security networks and global trade flows. These structural features make the country an important site for understanding how geopolitical narratives circulate and become entangled within local media systems.

Russia's engagement with Latin America must be understood within this broader geopolitical landscape. Scholars of Russian foreign policy and media strategy have shown how the Kremlin has sought to cultivate relationships with states that express ambivalence toward U.S. influence or embrace multipolar framings of global order (Benkler, Faris, and Roberts 2018; Pomerantsev 2014; Yablokov 2015). In this context, Spanish-language media outlets, such as *RT en Español* and *Sputnik Mundo*, play a central

https://doi.org/10.1017/chr.2026.10031 Published online by Cambridge University Press

role by providing regionally tailored content that challenges Western institutions, foregrounds critiques of U.S. foreign policy, and appeals to audiences with existing historical or ideological grievances (Brandt and Wirtschafter 2022; Rouvinski 2022). This media expansion reflects a broader effort to position Russia as an influential normative actor within the Global South, offering alternative understandings of international events and cultivating discursive alignment.

Within Colombia, these global dynamics intersect with domestic vulnerabilities, including political polarization, persistent institutional distrust, and longstanding conflicts involving state and non-state actors (Levitsky and Murillo 2013; Lupu, Ramírez Bustamante, and Zechmeister 2020). These conditions create openings for foreign narratives to be selectively adopted, reframed, or contested within local media coverage. While Russian outlets directly reach Colombian audiences through digital platforms and partnership arrangements (Rouvinski 2022), their potential influence also operates indirectly through the circulation of syndicated wire stories, regionally focused commentary, and the broader discursive space in which geopolitical meaning is negotiated.

By situating Colombia within these intersecting regional and global currents, this study examines not whether Russian media exert direct influence over Colombian discourse, but how narratives prominent in one media ecosystem may resonate with, align with, or be repurposed within another. This contextual framing lays the groundwork for analyzing narrative convergence without presuming intentional coordination or linear models of influence.

Mechanisms of Russian messaging

Russian engagement in the global information environment is multifaceted, combining state-backed media production, regionally tailored messaging, and strategic framing of international events. In Latin America, outlets, such as *RT en Español* and *Sputnik Mundo*, play a central role in this strategy by providing Spanish-language coverage that positions Russia as an alternative narrative authority and challenges the predominance of Western media flows (Oates 2016; Pomerantsev 2014; Yablokov 2015). Rather than functioning solely as propaganda organs, these outlets operate as hybrid media actors: they mix factual reporting with discursive framings, and they embed global narratives within regionally resonant contexts (Brandt and Wirtschafter 2022; Rouvinski 2022).

A growing body of scholarship shows that Russian international media employ several recurring narrative strategies. These include: (1) emphasizing Western hypocrisy or double standards in foreign policy; (2) foregrounding themes of multipolarity, sovereignty, and resistance to interventionism; (3) amplifying critiques of U.S. security partnerships; and (4) framing global conflicts through affective registers, such as crisis, escalation, or geopolitical danger (Pomerantsev 2019; Starbird, Arif, and Wilson 2019). These strategies do not operate only through overt persuasion, but through discursive positioning, offering discursive templates that audiences or local media may adopt, contest, or repurpose in their own contexts.

In Colombia, where institutional trust is comparatively low and geopolitical skepticism has deep historical roots (Levitsky and Murillo 2013; Lupu, Ramírez Bustamante, and Zechmeister 2020), such narrative strategies may find discursive resonance. Russian state-backed outlets reach Colombian audiences directly through digital distribution and, as documented by Rouvinski (2022), through media partnerships that expand the availability of Spanish-language broadcasts. Yet the ways narratives associated with Russian media circulate within Colombian discourse are not limited to direct consumption. Narratives may also diffuse indirectly via syndicated wire reporting, selective quoting of official statements, or the broader intertextual field in which geopolitical meaning is constructed (Benkler, Faris, and Roberts 2018; Oates 2016).

It is important to emphasize that the presence of frames associated with Russian media within Colombian coverage does not imply coordination, intent, or direct influence. Instead, this study conceptualizes Russian messaging as part of a discursive ecosystem, in which narratives can travel, be recontextualized, or intersect with pre-existing political and cultural orientations. In this view, "mechanisms" refer not to covert operational pathways but to media practices, stylistic framings, and narrative templates that structure how events are rhetorically presented and how meanings circulate across linguistic and geopolitical boundaries.

By focusing on these discursive mechanisms, rather than assuming intentional influence, this study examines how Russian narratives acquire local inflections when they appear in Colombian media, how they align with domestic grievances or ideological currents, and how they participate in the broader symbolic contestation of geopolitical events. This interpretive focus provides the grounding necessary for analyzing temporal narrative convergence without presupposing causal transmission or coordinated messaging campaigns.

Research gaps and objectives

Despite growing attention to Russian international media and transnational information flows, substantial gaps remain in understanding how narratives evolve, diffuse, and become recontextualized within local media ecosystems, particularly in Latin America. Much of the scholarship on influence operations focuses on platform-level dynamics (such as bot networks, audience analytics, or virality patterns) primarily within English-language or Euro-Atlantic contexts. These studies provide insight into the mechanics of content dissemination but often overlook the interpretive dimensions of how narratives are transformed as they move across linguistic, cultural, and geopolitical boundaries.

Within Latin America, research on Russian media strategy has been comparatively limited, with notable exceptions such as Rouvinski (2022), who documents the institutional infrastructure of Russian broadcast operations. Yet this work relies primarily on qualitative analysis and does not incorporate scalable computational methods capable of tracing narrative transformation across large multilingual corpora. Similarly, while recent advances in topic modeling, semantic clustering, and text embeddings offer promising tools for understanding narrative structure (Caldara and Iacoviello 2022; Grootendorst 2022; Wang et al. 2024), many applications remain focused on frequency-based signals or monolingual corpora, leaving unexplored the cross-lingual dynamics central to narrative circulation in Latin America.

A further gap concerns how cross-ecosystem narrative alignment can be identified without presuming direct transmission, coordinated campaigns, or unambiguous origin tracing. Existing studies often focus either on overt influence operations or on platform-level dissemination, leaving less room for analyses of discursive convergence detectable through patterns of semantic alignment, affective tone, and temporal sequencing. The present study therefore focuses on *narrative mimicry* and *discursive convergence*: observable similarities in framing, emphasis, and evaluative structure across media ecosystems, regardless of whether those similarities arise from direct uptake, shared event

https://doi.org/10.1017/chr.2026.10031 Published online by Cambridge University Press

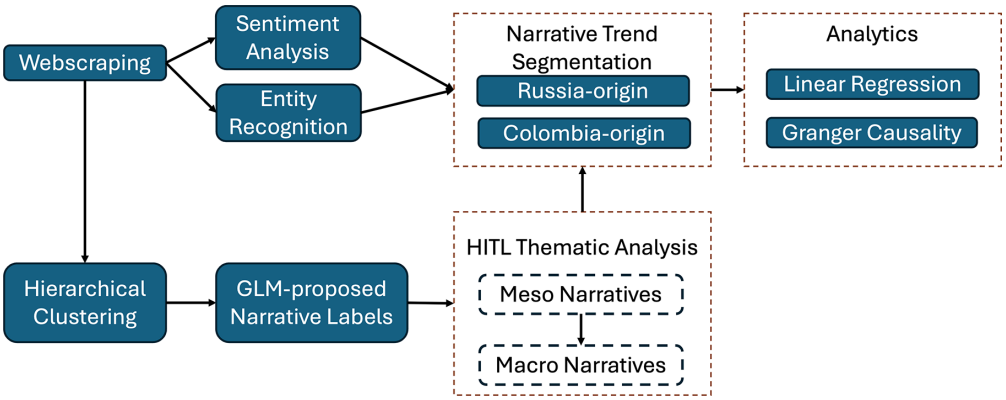


Figure 1. Methodological workflow integrating computational and interpretive stages.

exposure, journalistic routines, or broader ideological resonance. Addressing these dynamics requires combining computational text analysis with interpretive attention to meaning-making, context, and discursive resonance. In media studies, framing theory establishes that meaning in news texts is constructed through selective emphasis, causal attribution, and evaluative tone (Entman 1993, 2004) – analytical dimensions that computational methods can surface at scale but that require interpretive adjudication to render meaningful. Boumans and Trilling (2016) articulate this complementarity specifically for computational news analysis, arguing that automated content approaches gain validity through integration with domain-grounded qualitative frameworks. Recent work on cross-lingual narrative clustering further demonstrates how computational methods can identify structural similarities across media ecosystems while remaining dependent on human interpretation to establish their significance (Schneider et al. 2024). This study therefore follows the broader humanities data analysis tradition of combining computational procedures with domain qualitative context and case-based analytical reasoning, as articulated in textbook treatments of the field that emphasize iterative movement between quantitative pattern detection and interpretive close reading (Karsdorp, Riddell, and Kestemont 2021).

Colombia presents a particularly compelling site for this investigation. As a regionally influential U.S. ally with longstanding patterns of institutional distrust (Filippidou and O'Brien 2020; Shenk 2022) and a historically polarized media environment (Lupu, Ramírez Bustamante, and Zechmeister 2020), Colombia's information ecosystem offers fertile ground for understanding how geopolitical narratives are adopted, adapted, or contested. Russian media have increasingly targeted Spanish-speaking audiences (Brandt and Wirtschafter 2022), not simply by providing alternative viewpoints but by introducing narrative frames that intersect with local political grievances and symbolic vocabularies. Understanding how and why particular narratives resonate in this context requires methods sensitive not only to linguistic patterns but also to discursive context and temporal dynamics.

To address these gaps, this study pursues four interlinked objectives:

1. To identify recurring narrative structures in Russian state-backed media and examine how similar structures appear in Colombian coverage through patterns of discursive convergence and narrative mimicry.
2. To analyze sentiment alignment and affective framing across Russian and Colombian sources, with particular attention to narratives involving the United States and the 2014 and 2022 Russia–Ukraine conflicts.
3. To examine the role of intermediary agents, including regional outlets and syndicated wire services, in mediating the circulation, reframing, or domestication of narrative frames associated with Russian media.
4. To demonstrate how advanced language modeling, combined with semi-automated qualitative coding, can produce interpretable hierarchical narrative structures that illuminate the evolving semantic landscape of geopolitical discourse.

Taken together, these objectives reflect a methodological commitment to scalable interpretation, in which computational tools are used not merely to detect patterns in text, but to situate those patterns within a humanistic framework attentive to meaning, resonance, symbolic power, and the cultural life of narratives across borders.

Methodology

This study employs a modular, six-stage analytical workflow (Figure 1) designed to trace the emergence, convergence, and recontextualization of geopolitical narratives across Russian and Colombian media. The pipeline integrates multilingual preprocessing, NER, sentiment analysis, semantic embeddings, and hierarchical topic modeling, with interpretive adjudication provided through a structured HITL process. Throughout, the methodological design follows best practices for transparent annotation and collaborative interpretation as articulated in the INCEpTION framework (Klie et al. 2018) and aligns with FAIR data-management principles (Wilkinson et al. 2016). The workflow is intentionally hybrid: computational scale enables the detection of broad narrative patterns, while humanistic evaluation ensures that meaning, nuance, and discursive context remain central to the analysis.

Data collection

A total of 38,259 articles published between 1 January 2013 and 31 December 2023 were collected from six Colombian and four Russian outlets ("Selected sources" section in the Appendix). Articles and metadata were extracted using RSS feeds when available and scraped with Python's `BeautifulSoup` or Selenium when dynamic content loading required a headless browser. All documents were stored in a secure, on-premise MongoDB instance for reproducibility and auditability.

https://doi.org/10.1017/chr.2026.10031 Published online by Cambridge University Press

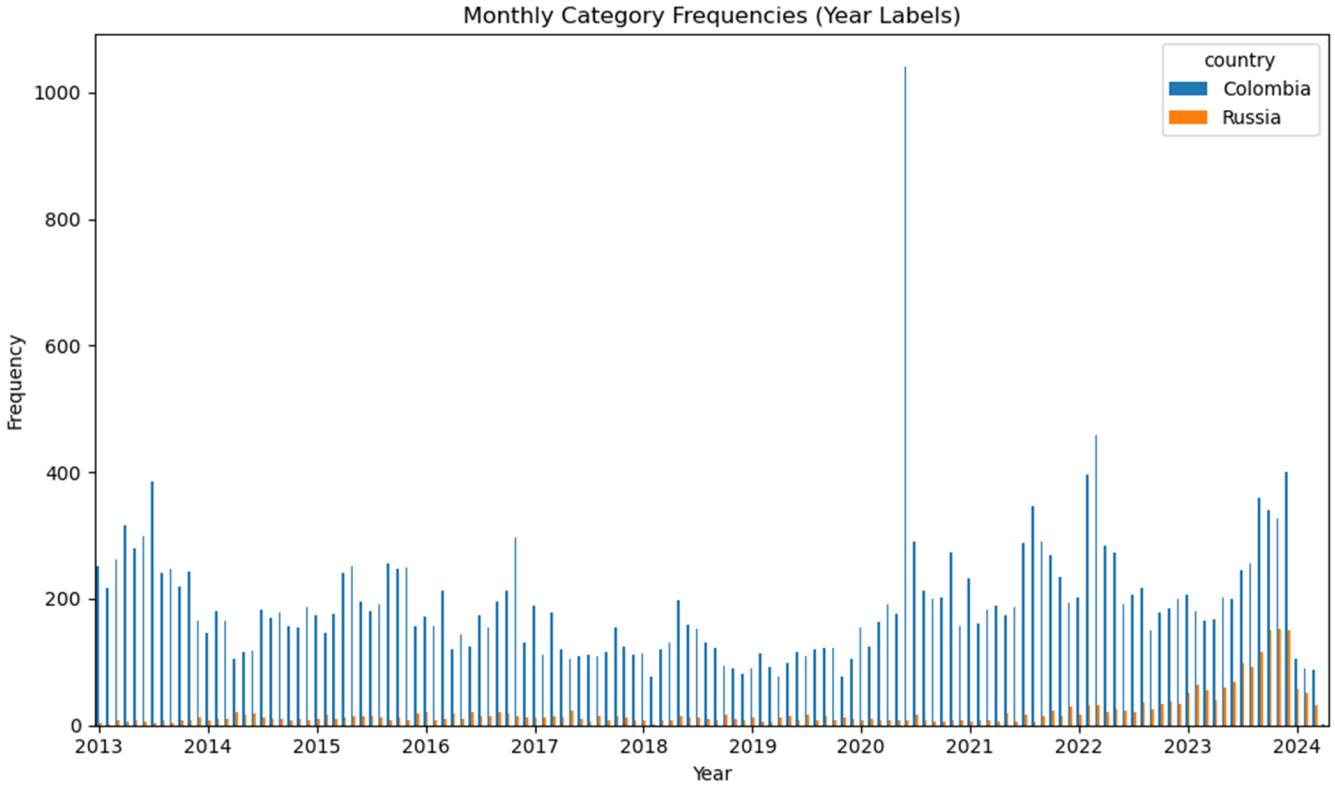


Figure 2. Raw count of monthly articles.

To ensure comprehensive coverage while minimizing topical bias, Colombian sources were retrieved using a multilingual query list focused on key geopolitical actors (e.g., *Rusia*, *Estados Unidos*, and *OTAN*). Following the audit framework of Olteanu et al. (2019), we drew a random sample of 5,535 texts (14% of the corpus) for dual human coding. Approximately 16 percent were deemed out-of-scope. Because this sample size yields a worst-case 95 percent margin of error of ±1 percentage point, we can generalize the off-topic rate to the remainder of the corpus with high confidence. During HITL aggregation, 98 percent of irrelevant documents were automatically routed to a dedicated "Miscellaneous/Off-topic" category, reducing semantic noise and preserving the validity of later clustering.

To provide a corpus-level sense of scale, Figures 2 and 3 plot monthly article counts for Russian and Colombian outlets. Raw counts illustrate the large volume disparity, while log-normalized counts expose relative temporal fluctuations essential for the later time-series analyses.

Named entity recognition and sentiment analysis

To connect narrative patterns to specific geopolitical referents, we extracted structured information from each article using multilingual NER and document-level sentiment analysis.

NER was conducted using the "Babelscape/wikineural-multilingual-ner" transformer model (Tedeschi et al. 2021), which achieves robust performance across Spanish and Russian benchmarks. Approximately 3,500 articles contained ambiguous or metonymic location references (such as capital cities used as stand-ins for nations). For these cases, a controlled toponymic disambiguation step was performed using a locally hosted Llama-2-7B-Nous-Hermes2 model. The model was supplied with tailored prompts and cross-checked with language identification tools (Lui and Baldwin 2012; Tiedemann and Ljubešić 2012). Manual validation of a 350-document subsample indicated approximately 99 percent accuracy, reflecting the high regularity of journalistic naming conventions. Country mappings were standardized using the `country_converter` library (Stadler 2017).

Sentiment was estimated with the multilingual BERT sentiment model (`nlptown/bert-base-multilingual-uncased-sentiment`), applied to overlapping text windows and averaged per document. Although transformer-based sentiment models perform well across languages (Aßenmacher, Corvonato, and Heumann 2021; Palomino and Ochoa-Luna 2020), we note that geopolitical discourse often relies on metaphor, euphemism, and pragmatic inference – features difficult for automated systems to fully capture. The sentiment layer is therefore treated as a coarse but useful indicator of evaluative tone. Future analyses could incorporate aspect-based sentiment analysis (Pontiki et al. 2016) to tie sentiment more precisely to actors or events.

Topic modeling and semantic embeddings

Topic modeling was used to identify coherent thematic structures and trace their evolution across time, sources, and languages ("Topic modeling definitions" section in the Appendix). We employed the `intfloat/multilingual-e5-large-instruct` embedding model, chosen for its demonstrated performance on English–Spanish–Russian corpora (Wang et al. 2024). Embeddings were clustered using BERTopic (Grootendorst 2022), which combines UMAP for dimensionality reduction (McInnes, Healy, and Melville 2018), agglomerative clustering for semantic grouping, and class-based TF-IDF (CTFIDF) for substantive keyword extraction.

https://doi.org/10.1017/chr.2026.10031 Published online by Cambridge University Press

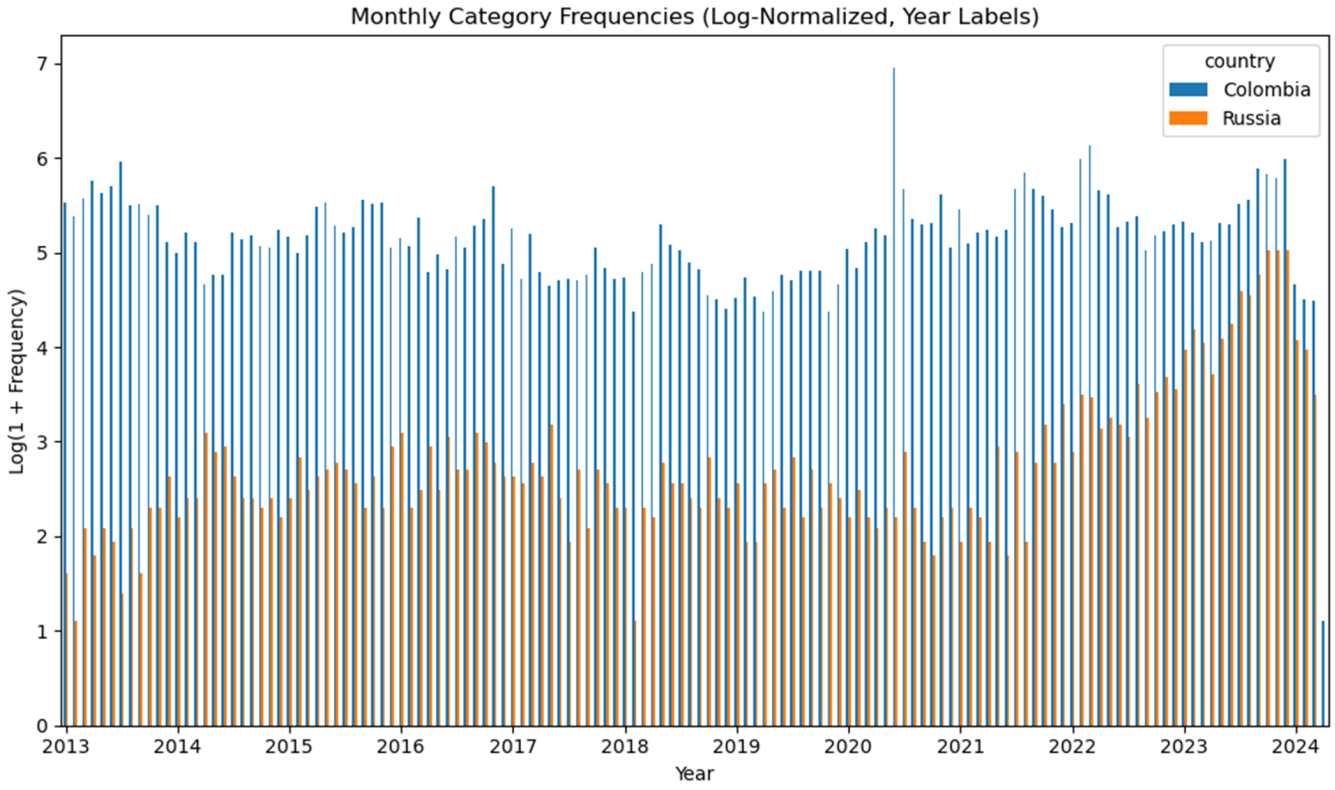


Figure 3. Log normalized count of monthly articles.

Cluster granularity was tuned through iterative experimentation informed by the Calinski–Harabasz metric via the `KElbowVisualizer` toolkit (Bengfort and Bilbro 2019). Comparative tests with HDBSCAN showed that agglomerative clustering provided more stable and semantically coherent clusters for this corpus, particularly in avoiding large, diffuse "catch-all" groups common in multilingual settings.

Representative documents and cluster keywords were then used to generate candidate labels via a locally hosted Llama-2-7B-Nous-Hermes2 model. This label-suggestion step follows recent work showing the utility of LLMs in qualitative classification tasks (Meng et al. 2026; Tai et al. 2024). The model's role was explicitly heuristic: it produced initial hypotheses about cluster meaning, which were subsequently refined or replaced by human coders.

Cluster labeling and narrative hierarchies

To transform computational clusters into analytically meaningful categories, we implemented a structured HITL labeling protocol rooted in qualitative coding methods ("Cluster label candidates prompt example" section in the Appendix). For each cluster, three coders independently reviewed representative texts, LLM-generated candidate labels, and CTFIDF keywords. Coders applied a three-stage procedure – initial, focused, and axial coding (Saldaña 2021) to select or refine the label that best captured the cluster's semantic core. Consensus meetings resolved disagreements and ensured thematic coherence, linguistic accuracy, and domain relevance. This process mirrors best practices in collaborative human–AI interpretive workflows (Feuston and Brubaker 2021; Hoyle et al. 2021; Jiang et al. 2021).

Clusters were then aggregated into a three-level narrative hierarchy:

1. **Micro-narratives:** Individual clusters reflecting specific events or discourse fragments.
2. **Meso-narratives:** Groups of micro-narratives forming broader thematic arcs (such as "Russia–Ukraine conflict").
3. **Macro-narratives:** High-level conceptual domains, such as diplomacy, security & conflict, economy, and politics & society.

Aggregation was guided by principles of thematic coherence, distinctiveness, and narrative evolution. Coders validated relationships across levels by cross-checking representative keywords, cluster summaries, and temporal characteristics. The resulting hierarchy provides the conceptual backbone for analyzing narrative resonance, convergence, and recontextualization across media ecosystems.

Limitations and challenges

Several methodological limitations warrant consideration. First, multilingual preprocessing introduces challenges related to idiomatic variation, indirect sentiment, and metonymic reference. Although native Spanish- and Russian-speaking reviewers validated a large subsample, nuances in evaluative language may elude automated classifiers. Second, reliance on state-backed media introduces a structural bias: these sources intentionally reflect institutional priorities. While cross-source triangulation mitigates this concern, narrative patterns may still reflect editorial agendas rather than broader discursive currents.

Third, the hierarchical aggregation of clusters into narratives involves judgement. Although consensus coding, structured protocols, and LLM-assisted ideation reduce subjectivity, some ambiguity is inherent in narrative categorization. Finally, topic models impose static boundaries on narratives that are often fluid. Future

https://doi.org/10.1017/chr.2026.10031 Published online by Cambridge University Press

Table 1. Descriptive statistics for significant weekly regressions, 2022–2023 ($p < 0.05$)

Macro-narrative	Min	Max	Mean	Dom. Lag (wks)	N_{sig} (raw)	N_{sig} (BH)
Economy	0.35	0.88	0.57	2.8	8	5
Security and conflict	0.34	0.73	0.51	1.0	12	12
Diplomacy	0.42	0.76	0.55	2.2	5	4
Politics and society	0.48	0.92	0.67	2.4	7	4

Note: Values represent absolute Pearson correlation coefficients at the dominant lag for significant tests ($p < 0.05$). N_{sig} (raw) indicates tests with $p < 0.05$ prior to Benjamini–Hochberg adjustment; N_{sig} (BH) indicates tests surviving FDR correction. *Politics and Society* totals reflect 8 available combinations (vs. 12 for other narratives) owing to missing target–year cells in 2022.

research could apply dynamic topic modeling (Blei and Lafferty 2006) or streaming embeddings to capture shifting semantic fields more precisely.

Despite these challenges, the combined use of computational scale, linguistic expertise, and structured adjudication provides a robust methodological foundation for analyzing how narrative frames associated with Russian media align with and interact with Colombian media ecosystems.

Results

This section reports the quantitative patterns identified through weekly aggregation of narrative frequency and sentiment, followed by an illustrative case study. In keeping with the cautions articulated earlier, these results should be read as indicators of temporal association and narrative convergence rather than as evidence of direct causal influence.

Temporal patterns in narrative frequency and sentiment

After normalizing narrative frequency (0–1) and sentiment (−1 to 1), we estimated a series of time-trend models and lagged cross-series tests, including Granger causality models, across macro-narratives and geopolitical targets. The full dataset yielded 330 narrative–target combinations; across five model specifications, this produced 1,650 regressions. For interpretability, and to align with the study's geopolitical focus, we restricted our analysis to the two log-normalized models and to periods surrounding the 2014 and 2022 Russia–Ukraine conflicts, yielding 360 tests (180 per conflict period).

The 2014 results were generally inconsistent across years and narrative domains. This is not surprising given the limited penetration of Russian Spanish-language media during that period and the smaller volume of Colombian reporting on the early stages of the Ukraine crisis. As a result, the 2014 patterns do not allow confident explanation and are not discussed further. Future work incorporating subnational Colombian media or social-media content may increase signal clarity for this earlier period.

By contrast, analyses surrounding the 2022 conflict revealed more coherent temporal relationships. As a descriptive reference point, we examined categories with comparatively low geopolitical salience (e.g., *Technology*, *Health*, and *Irrelevant*). These categories displayed less consistent lag structures than the geopolitically focused narratives discussed below, though we caution that this comparison does not constitute a formal control: any major geopolitical event can generate synchronized coverage across media systems through shared event exposure alone, and some ostensibly non-geopolitical categories (such as health) intersect with documented Russian messaging strategies under specific conditions. The contrast with these categories is therefore offered as context rather than as evidence of uniqueness.

Against this descriptive backdrop, four macro-narratives, *Economy*, *Security and Conflict*, *Diplomacy*, and *Politics and Society*, displayed patterns of temporal coupling between Russian and Colombian coverage across the 2022–2023 period ("Macro-narratives" section in the Appendix). Within this focused set of 44 tests (4 narratives × 3 geopolitical targets × 2 metrics × 2 years, with *Politics and Society* represented by 8 rather than 12 combinations owing to limited Russian and Ukrainian entity presence in Colombian political coverage during 2022), 32 met the $p < 0.05$ threshold prior to false-discovery correction, with 25 surviving Benjamini–Hochberg adjustment. The strength and consistency of these patterns varied considerably across narratives. *Security and Conflict* exhibited the most robust temporal alignment, with all 12 tests surviving FDR correction across both years and all three geopolitical targets. *Economy*, *Diplomacy*, and *Politics and Society* showed more variable patterns, with post-BH retention rates of five, four, and four tests, respectively; in all three narratives, surviving tests were concentrated primarily in sentiment rather than frequency alignment. Table 1 reports descriptive characteristics for each narrative category.

Across these narrative bundles, a recurring temporal structure emerged, most consistently within *Security and Conflict* and, to a lesser degree, *Economy*, *Diplomacy*, and *Politics and Society*: in many cases, Russian narrative volume predicted Colombian volume most strongly at lag 1, while sentiment alignment, when present, appeared one additional week later. This sequential pattern, in which coverage volume intensifies prior to evaluative convergence, aligns with media agenda-setting research suggesting that salience cues often precede tonal framing (Entman 2004; McCombs and Shaw 1972; Vliegenthart and Walgrave 2008). Importantly, Granger causality here indicates only temporal precedence. It does not establish intention, coordination, or a unidirectional flow of influence. The observed lag ordering nevertheless identifies temporal windows of narrative alignment that warrant closer qualitative inspection.

Based on this broader pattern, we selected one narrative–target pair – *USA* within the *Security and Conflict* macro-narrative (2023) – as an illustrative case. This bundle was chosen because it (i) relates directly to the study's geopolitical focus, (ii) exhibits statistically robust lag structures, and (iii) maintains a context-sensitive temporal trajectory across both frequency and sentiment. It is presented not as a representative or exceptional case, but as a clear instance of the types of patterns, the methodology can surface.

Figure 4 shows weekly log-normalized coverage volume. Russian outlets (red) consistently exhibit greater intensity than Colombian outlets (blue), with both series rising and falling in tandem ($r = 0.688$, $R^2 = 0.47$). Granger tests indicate that Russian coverage

https://doi.org/10.1017/chr.2026.10031 Published online by Cambridge University Press

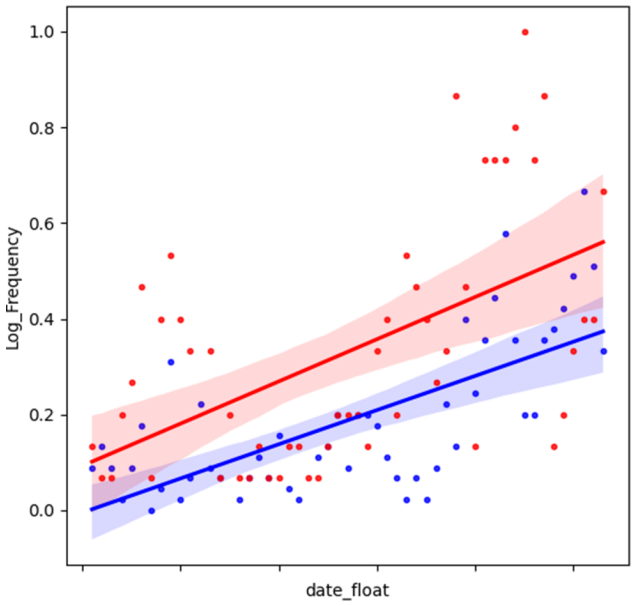


Figure 4. 2023 normalized frequency of USA mentions within security and conflict narrative.

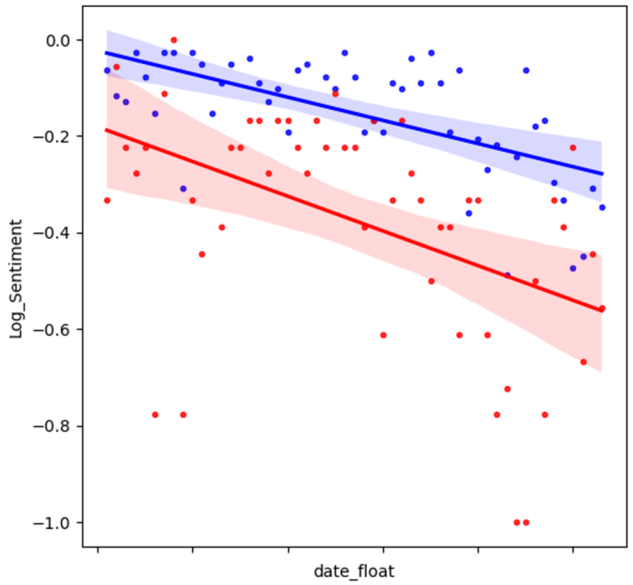


Figure 5. 2023 normalized sentiment of USA mentions in security and conflict narrative.

significantly predicts Colombian coverage at lag 1 ($p = 0.032$) and marginally at lag 2 ($p = 0.056$).

Figure 5 displays sentiment trajectories ($r = 0.613$, $R^2 = 0.38$). Sentiment is negative in absolute terms for both media ecosystems, and Granger tests suggest a weaker but still substantive alignment (lag 1: $p = 0.095$; lag 2: $p = 0.050$), consistent with the sequential salience-first, tone-second pattern observed across the dataset.

While this case is highlighted for interpretive clarity, similar lag structures – volume alignment at lag 1 and sentiment alignment at lag 2 – appear throughout the 2022–2023 subset. Such patterns are consistent with a general tendency for Colombian coverage to increase shortly after spikes in Russian reporting on salient geopolitical issues, followed by gradual convergence in evaluative framing, although these relationships may also reflect shared responses to major geopolitical events rather than direct transmission.

Case study: Convergent negative framing of Western arms deliveries

To complement the quantitative signals above, this section provides an illustrative micro-case that demonstrates how temporally adjacent articles can exhibit shared evaluative framing. This example is intended to show what the methodology can reveal; it is not offered as evidence of direct narrative transfer or coordinated influence.

Using the meso-narrative hierarchy developed earlier, we identified 30 Russia–Colombia document pairs within the *Ukraine Conflict* meso-narrative in which the Russian article preceded the Colombian one by less than two weeks. We then applied a bilingual lexicon-based tonality measure (e.g., *guerra*, *crisis*, *agresión*, *escalada*, and *paz*) and manually inspected cases with high negative scores.

One pair stood out for its thematic and tonal similarity. The first article, published by Russian state media in late January 2023, features commentary by analyst Enrique Refoyo and frames Western weapons deliveries as escalatory, cynical, and rooted in geopolitical self-interest. It casts Western states as instrumentalizing Ukraine for broader strategic aims and downplays Ukrainian agency, a framing consistent with long-standing Russian narratives about Western destabilization.

Approximately eight days later, a Colombian article in *El Universal*, drawing on Deutsche Welle (DW) reporting, foregrounded Russian warnings regarding long-range missile deliveries. The piece emphasized that such weapons could strike "todo el territorio controlado por los rusos" and cited Kremlin spokesperson Dmitri Peskov's claim that deliveries would cause a "escalada significativa del conflicto." Although the Colombian article attributes these claims to Russian officials rather than adopting them as editorial positions, its core framing similarly emphasizes risk, escalation, and the destabilizing consequences of Western decisions.

Viewed together, the pair illustrates a form of discursive convergence: two media systems describing the same event within a similar evaluative frame, oriented around danger, escalation, and Western responsibility. In this example, the Colombian article relies on wire-service material that itself foregrounds Russian discourse, illustrating how circulation channels can mediate the movement of narrative frames without requiring direct coordination.

This case is not presented as proof of influence or intentional messaging transfer. Rather, it demonstrates the kind of cross-ecosystem alignment in timing, negative tone, and causal framing that the methodological framework is designed to detect. It also underscores the interpretive importance of examining not only statistical associations but also how narratives travel, are reframed, and acquire meaning within specific media ecologies.

Discussion

This study demonstrates how combining large-scale computational text analysis with an interpretive narrative framework can illuminate the dynamics through which geopolitical narratives emerge, converge, and circulate across distinct media ecosystems. Rather than seeking direct evidence of coordination or causal influence, our goal has been to identify patterned relationships in narrative salience and evaluative framing that become visible only when multilingual corpora are examined at scale. The results suggest that certain domains, especially those related to security, diplomacy,

https://doi.org/10.1017/chr.2026.10031 Published online by Cambridge University Press

and U.S.–Russia relations, exhibit recurring temporal sequences in which spikes in Russian coverage are followed by shifts in Colombian narrative volume and, to a lesser degree, sentiment. These associations do not constitute proof of narrative transfer but do represent conditions under which convergence becomes analytically meaningful.

At the narrative level, the study's hierarchical framework clarifies how similarities across media ecosystems can arise not merely from shared vocabulary but from alignment in the meso-level narrative structures through which geopolitical events are contextualized. The patterns observed during the 2022 Russia–Ukraine conflict period indicate that Colombian coverage sometimes reflects frames emphasizing escalation, destabilization, or Western responsibility that are also prominent in Russian state media. In several cases, this alignment appears mediated through widely used international wire services, demonstrating how circulation channels can transmit evaluative cues even without direct ideological affinity. The qualitative case study further illustrates how two outlets, writing eight days apart, foreground similar causal claims about the risks of Western weapons transfers, albeit with different degrees of editorial endorsement.

These results align with prior research on intermedia agenda-setting (McCombs and Shaw 1972; Vliegenthart and Walgrave 2008) and convergence in transnational news flows. They support the view that narrative mimicry is best understood not as the replication of content, but as the repetition of framing logics, metaphors, causal attributions, and tonal structures. In this sense, the present study contributes to computational humanities by demonstrating how large-scale textual patterns can be situated within substantive questions about meaning, resonance, and discursive power. It underscores that computational approaches can reveal regularities in narrative architecture that invite, rather than replace, humanistic evaluation.

At the same time, the findings call attention to the layered nature of narrative movement. Some alignments likely reflect structural media routines (like reliance on shared newswires), while others may reflect ideological consonance or strategic amplification. Distinguishing among these mechanisms requires integrating computational signal with qualitative interpretation, archival context, and media-production knowledge – an interdisciplinary task central to the aims of computational humanities.

Limitations

Several limitations qualify these interpretations. First, the corpus is restricted to open-access national outlets and government press releases, which omits paywalled publications, subnational media, and social media platforms where narrative dynamics may differ substantially. Inclusion of these sources could reveal either stronger echoing patterns or countervailing local framings that are not captured here.

Second, sentiment analysis was performed at the document level and therefore cannot distinguish between sentiment directed toward different actors within the same article. While the large scale of the dataset mitigates some noise, future work could incorporate aspect-based or entity-linked sentiment to isolate how specific countries or actors are evaluated within a narrative.

Third, topic modeling and hierarchical clustering impose discrete boundaries on narratives that, in practice, are fluid and overlapping. Although the HITL adjudication process reduced misclassification and ensured contextual coherence, some degree of interpretive subjectivity is unavoidable. This is a feature, rather than a flaw, of narrative analysis: narratives are culturally and symbolically constructed objects that resist purely algorithmic classification.

Fourth, the temporal analyses rely on linear Granger causality models, which provide information about predictive precedence but cannot adjudicate among alternative explanations, such as shared external shocks, global news cycles, or parallel editorial decisions. The observed lag structures should therefore be understood as indicative patterns that prompt further inquiry, not evidence of directional influence.

The negative-control categories (Technology, Health, and Off-topic) require particular interpretive caution. Their relative lack of temporal coupling is consistent with the view that the strongest alignments occur in geopolitically salient domains. However, this reasoning holds only if those categories fall genuinely outside Russian messaging priorities – an assumption rather than a finding. Health, for instance, intersects with documented Russian narratives around COVID-19 vaccine skepticism, meaning the controls may not be entirely clean. More fundamentally, geopolitical crises generate synchronized coverage across many country pairs through shared event exposure alone; coupling in security and diplomacy narratives might be expected between *any* two media systems covering the same international events, not only between Russian-backed and Colombian outlets. Without a comparative baseline drawn from an unrelated country pair covering the same period, the design cannot fully disentangle media-ecosystem-specific alignment from general event-driven synchronization.

The 2014 conflict period offers a partial within-study temporal reference point. The absence of consistent lag structures during that earlier period, when Russian Spanish-language media infrastructure was comparatively less developed (Rouvinski 2022), is at least consistent with the view that the 2022 patterns are not a simple artifact of long-run structural correlation between these two media systems. This comparison cannot serve as a clean control, however: the marked difference in Russian media reach between 2014 and 2022 is itself a confound, and the contrast in findings across periods reflects both the maturation of that infrastructure and any genuine shift in narrative alignment. These constraints underscore the importance of treating all temporal associations as suggestive and hypothesis-generating rather than confirmatory.

Finally, multilingual and cross-genre corpora introduce challenges related to idiomatic variation, translation effects, and geopolitical framing differences across languages. While native-speaker validation helped mitigate these issues, subtleties in evaluative language may nonetheless escape automated detection.

Future research directions

Future research should expand both the conceptual and methodological scope of the present framework. Incorporating social media data, regional or local Colombian outlets, and paywalled journalism would allow for a more fine-grained understanding of how narratives diffuse across institutional boundaries and audience segments. Comparative work could examine whether similar patterns of convergence arise in other regions targeted by Russian, Chinese, or U.S. state media, enabling broader theorization of narrative circulation under conditions of geopolitical competition. Finally, future work should more fully explore the intermediaries (wire services, syndication chains, diplomatic channels, and influencers) through which evaluative frames move across media ecosystems. Network-analytic approaches could help trace these circulation patterns explicitly, allowing for a richer

https://doi.org/10.1017/chr.2026.10031 Published online by Cambridge University Press

account of how narratives are transformed as they travel. The present study operationalized this objective primarily through a single illustrative case; systematic intermediary tracing across the full corpus remains an important open task.

Methodologically, dynamic topic modeling or nonlinear approaches (VAR models with threshold effects and attention-based temporal architectures) may better capture complex narrative trajectories that linear models oversimplify. Integrating entity-resolution pipelines could support actor-specific sentiment or stance analysis, offering deeper insight into how narratives attach evaluative meaning to particular geopolitical subjects.

Taken together, these directions underscore the promise of integrating computational scale with humanities-oriented analysis. By providing tools for identifying when and how discursive alignments occur, this study invites further inquiry into the cultural, institutional, and geopolitical conditions that shape the global life of narratives.

Conclusions

This study demonstrates how computational text analysis, when combined with an interpretive narrative framework, can illuminate patterned relationships in the ways geopolitical narratives emerge, circulate, and align across media ecosystems. By examining more than a decade of Russian and Colombian news coverage, and by situating large-scale linguistic patterns within a structured hierarchy of micro-, meso-, and macro-narratives, the analysis reveals recurring temporal and thematic convergences particularly in domains related to security, diplomacy, and U.S.–Russia relations.

These findings do not establish direct influence or coordinated information flows. Rather, they highlight the conditions under which narrative mimicry, understood here as alignment in framing logics, evaluative tone, and discursive structure, becomes analytically visible. The observed sequences in which Russian narrative salience often precedes similar shifts in Colombian coverage suggest points of contact that merit closer qualitative attention. The illustrative case study further shows how evaluative frames can converge across outlets even in the absence of shared editorial intent, sometimes mediated by international wire services or common global news cycles.

Methodologically, the study contributes to computational humanities by demonstrating how multilingual embeddings, hierarchical topic modeling, and HITL adjudication can be integrated into a unified interpretive workflow. This hybrid approach bridges computational scale and qualitative nuance, enabling the analysis of narrative architecture at a level of granularity and volume that would be difficult to achieve through manual methods alone. It also offers a transparent and reproducible framework for identifying when, where, and how discursive alignments emerge across different media spheres.

The results underscore both the promise and the limitations of computational approaches to narrative analysis. While automated methods can surface patterns that warrant attention, their interpretive value ultimately depends on contextual grounding, human judgment, and an awareness of the cultural and geopolitical dynamics that shape textual production. The narrative convergences identified here therefore serve as starting points for deeper inquiry into how geopolitical frames travel, how they are refracted through local media logics, and how meaning is negotiated across linguistic and regional boundaries.

Taken together, this study highlights the value of interdisciplinary approaches that combine computational modeling with humanistic interpretation. As global information environments become increasingly interconnected and as state and non-state actors seek to shape them, such approaches offer a promising path for understanding the circulation, transformation, and contested life of narratives across the contemporary media landscape.

Data availability statement. Supplementary material and replication code are available at https://doi.org/10.5281/zenodo.19186851. Aggregated weekly narrative and Granger causality results are available as supplementary data. The underlying corpus cannot be redistributed due to copyright restrictions on scraped news content; access to article-level data may be requested from the corresponding author.

Acknowledgements. The authors thank the anonymous reviewers for their helpful comments, as well as colleagues and peers at George Washington University, Florida International University, the University of Virginia, and Elder Research for valuable discussions and feedback. All errors remain our own. J.P.'s work was conducted during his doctoral studies at Florida International University. This article builds on a Master's thesis submitted at George Washington University.

Author contribution. Conceptualization: J.P. and A.J.; Software: J.P.; Visualization: J.P.; Writing and reviewing code and paper: J.P. and A.J. Both authors approved the final submitted draft.

Funding statement. The authors declare that no specific funding has been received for this article.

Competing interests. The authors declare that no competing interests exist.

Ethical standard. The authors affirm that this research did not involve human participants.

References

Aßenmacher, Matthias, Alessandra Corvonato, and Christian Heumann. 2021. "Re-Evaluating GermEval17 Using German Pre-Trained Language Models." *CoRR* abs/2102.12330. https://arxiv.org/abs/2102.12330.

Baden, Christian, Chantal Pipal, Martijn Schoonvelde, and Mariken A. C. G. van der Velden. 2022. "Three Gaps in Computational Text Analysis for Political Science: Language, Validity, and Replication." *Political Communication* 16, no. 1: 1–18.

Bengfort, Benjamin, and Rebecca Bilbro. 2019. "Yellowbrick: Visualizing the Scikit-Learn Model Selection Process." *Journal of Open Source Software* 4, no. 35: 1075. https://doi.org/10.21105/joss.01075.

Benkler, Yochai, Robert Faris, and Hal Roberts. 2018. *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics*. New York, NY: Oxford University Press.

Blei, David M., and John D. Lafferty. 2006. "Dynamic Topic Models." In *Proceedings of ICML 2006*, 113–20. New York, NY: Association for Computing Machinery. https://doi.org/10.1145/1143844.1143859.

Boumans, Jelle W., and Damian Trilling. 2016. "Taking Stock of the Toolkit: An Overview of Relevant Automated Content Analysis Approaches and Techniques for Digital Journalism Scholars." *Digital Journalism* 4, no. 1: 8–23.

Brandt, Jessica, and Valerie Wirtschafter. 2022. "Working the Western Hemisphere: How Russia Spreads Propaganda about Ukraine in Latin America and the Impact of Platform Responses." Foreign policy at Brookings paper, Brookings Institution. https://www.brookings.edu/wp-content/uploads/2022/12/FP_20221216_russia_propaganda_brandt_wirtschafter.pdf.

Caldara, Dario, and Matteo Iacoviello. 2022. "Measuring Geopolitical Risk." *American Economic Review* 112: 4. https://doi.org/10.1257/aer.20191823.

Entman, Robert M. 1993. "Framing: Toward Clarification of a Fractured Paradigm." *Journal of Communication* 43, no. 4: 51–8.

Entman, Robert M. 2004. *Projections of Power: Framing News, Public Opinion, and U.S. Foreign Policy*. Chicago, IL: University of Chicago Press.

Feuston, Jessica L., and Jed R. Brubaker. 2021. "Putting Tools in their Place: The Role of Time and Perspective in Human-AI Collaboration for

Qualitative Analysis." *Proceedings of the ACM on Human-Computer Interaction* 5, no. CSCW2. https://doi.org/10.1145/3479856.

Filippidou, Anastasia, and Thomas O'Brien. 2020. "Trust and Distrust in the Resolution of Protracted Social Conflicts: The Case of Colombia." *Behavioral Sciences of Terrorism and Political Aggression* 14, no. 1. https://doi.org/10.1080/19434472.2020.1785524.

Gobat, Michel. 2013. "The Invention of Latin America: A Transnational History of Anti-Imperialism, Democracy, and Race." *The American Historical Review* 118, no. 5: 1345–75. https://doi.org/10.1093/ahr/118.5.1345.

Grimmer, Justin, Margaret E. Roberts, and Brandon M. Stewart. 2022. *Text as Data: A New Framework for Machine Learning and the Social Sciences*. Princeton, NJ: Princeton University Press.

Grootendorst, Maarten. 2022. "BERTopic: Neural Topic Modeling with a Class-Based TF-IDF Procedure." Preprint. https://arxiv.org/abs/2203.05794.

Hoyle, Alexander, Pranav Goel, Denis Peskov, Andrew Hian-Cheong, Jordan Boyd-Graber, and Philip Resnik. 2021. "Is Automated Topic Model Evaluation Broken?: The Incoherence of Coherence." In *Neural Information Processing Systems*. http://cs.umd.edu/~jbg//docs/2021_neurips_incoherence.pdf.

Jiang, Jialun Aaron, Kandrea Wade, Casey Fiesler, and Jed R. Brubaker. 2021. "Supporting Serendipity: Opportunities and Challenges for Human-AI Collaboration in Qualitative Analysis." *Proceedings of the ACM on Human-Computer Interaction* 5, no. CSCW1. https://doi.org/10.1145/3449168.

Jockers, Matthew L. 2013. *Macroanalysis: Digital Methods and Literary History*. Champaign, IL: University of Illinois Press.

Karsdorp, Folgert, Allen Riddell, and Mike Kestemont. 2021. *Humanities Data Analysis*. Princeton, NJ: Princeton University Press.

Klie, Jan-Christoph, Marine Bugert, Barbara Boullosa, Roland Eckart, and Iryna Gurevych. 2018. "INCEpTION: Towards a Platform for Machine-Assisted and Knowledge-Oriented Interactive Annotation." In *Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations*, Santa Fe, 5–9. Association for Computational Linguistics. https://aclanthology.org/C18-2002/.

Levitsky, Steven, and María Victoria Murillo. 2013. "Lessons from Latin America: Building Institutions on Weak Foundations." *Journal of Democracy* 24, no. 2: 93–107. https://doi.org/10.1353/jod.2013.0031.

Lui, Marco, and Timothy Baldwin. 2012. "Langid.py: An off-the-Shelf Language Identification Tool." In *Proceedings of ACL 2012, System Demonstrations*, edited by Min Zhang, 25–30. Jeju Island: Association for Computational Linguistics. https://aclanthology.org/P12-3005/.

Lupu, Noam, Mariana V. Ramírez Bustamante, and Elizabeth J. Zechmeister. 2020. "Social Media Disruption: Messaging Mistrust in Latin America." *Journal of Democracy* 31, no. 3: 160–71. https://doi.org/10.1353/jod.2020.0038.

McCombs, Maxwell E., and Donald L. Shaw. 1972. "The Agenda-Setting Function of Mass Media*." *Public Opinion Quarterly* 36, no. 2: 176–87. https://doi.org/10.1086/267990.

McInnes, Leland, John Healy, and James Melville. 2018. "UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction." *Journal of Open Source Software 3*, no. 29: 861. https://doi.org/10.21105/joss.00861.

Meng, Han, Yitian Yang, Fu Wayne, Jungup Lee, Yunan Li, and Yi-Chieh Lee. 2026. "Exploring the Human-LLM Synergy in Advancing Theory-Driven Qualitative Analysis." *ACM Transactions on Computer-Human Interaction* 33, no. 1. https://doi.org/10.1145/3778354.

Oates, Sarah. 2016. "Russian Media in the Digital Age: Propaganda Rewired." *Russian Politics* 1, no. 4: 398–417.

Olteanu, Alexandra, Carlos Castillo, Fernando Diaz, and Emre Kıcıman. 2019. "Social Data: Biases, Methodological Pitfalls, and Ethical Boundaries." *Frontiers in Big Data* 2: 13. https://doi.org/10.3389/fdata.2019.00013.

Palomino, Daniel, and José Ochoa-Luna. 2020. "Palomino-Ochoa at SemEval-2020 Task 9: Robust System Based on Transformer for Code-Mixed Sentiment Classification." In *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, edited by **Aurelie Herbelot, Xiaodan Zhu, Alexis Palmer, Nathan Schneider, Jonathan May, and Ekaterina Shutova**, 963–7. Barcelona: International Committee for Computational Linguistics. https://doi.org/10.18653/v1/2020.semeval-1.124.

Piper, Andrew. 2018. *Enumerations: Data and Literary Study*. Chicago, IL: University of Chicago Press.

Pomerantsev, Peter. 2014. *Nothing Is True and Everything Is Possible: The Surreal Heart of the New Russia*. New York, NY: PublicAffairs.

Pomerantsev, Peter. 2019. *This Is Not Propaganda: Adventures in the War against Reality*. New York, NY: PublicAffairs.

Pontiki, Maria, Dimitris Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Mohammad AL-Smadi, Mahmoud Al-Ayyoub, et al. 2016. "SemEval-2016 Task 5: Aspect Based Sentiment Analysis." In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, edited by **Steven Bethard, Marine Carpuat, Daniel Cer, David Jurgens, Preslav Nakov, and Torsten Zesch**, 19–30. San Diego, CA: Association for Computational Linguistics. https://doi.org/10.18653/v1/S16-1002.

Rouvinski, Vladimir. 2022. "The Misleading Truths of Russia's Strategic Communication in Latin America." *Global Security Review*, 2, no. 5. https://digitalcommons.fiu.edu/cgi/viewcontent.cgi?article=1012&context=gsr.

Saldaña, Johnny. 2021. *The Coding Manual for Qualitative Researchers*, 4th ed. Thousand Oaks, CA: SAGE.

Schneider, Nicole R., Kent O'Sullivan, Avik Das, and Hanan Samet. 2024. "Cross-Lingual Narrative Clustering Using Large Language Models." In *7th ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery (GeoAI'24)*, October 29-November 1, 2024, Atlanta, GA. New York, NY: Association for Computing Machinery. https://doi.org/10.1145/3687123.3698280.

Shenk, J. 2022. "Does Conflict Experience Affect Participatory Democracy after War? Evidence from Colombia." *Journal of Peace Research* 60, no. 6. https://doi.org/10.1177/00223433221105112.

Stadler, Konstantin. 2017. "The Country Converter Coco—A Python Package for Converting Country Names between Different Classification Schemes." *Journal of Open Source Software* 2, no. 16: 332. https://doi.org/10.21105/joss.00332.

Starbird, Kate, Ahmer Arif, and Tom Wilson. 2019. "Disinformation as Collaborative Work: Surfacing the Participatory Nature of Strategic Information Operations." In *Proceedings of the ACM on Human-Computer Interaction*, Volume 3, Issue CSCW. New York, NY: Association for Computing Machinery. https://doi.org/10.1145/3359229.

Tai, Robert H., Lillian R. Bentley, Xin Xia, Jason M. Sitt, Sarah C. Fankhauser, Ana M. Chicas-Mosier, and Barnas G. Monteith. 2024. "An Examination of the Use of Large Language Models to Aid Analysis of Textual Data." *International Journal of Qualitative Methods* 23. https://doi.org/10.1177/16094069241231168.

Tedeschi, Simone, Valentino Maiorca, Niccolò Campolungo, Francesco Cecconi, and Roberto Navigli. 2021. "WikiNEuRal: Combined Neural and Knowledge-Based Silver Data Creation for Multilingual NER." In *Findings of the Association for Computational Linguistics: EMNLP 2021*, edited by **Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-Tau Yih**, 2521–33, Punta Cana: Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.findings-emnlp.215.

Tiedemann, Jörg, and Nikola Ljubešić. 2012. "Efficient Discrimination between Closely Related Languages." In *Proceedings of COLING 2012*, edited by **Martin Kay and Christian Boitet**, 2619–34. Mumbai: The COLING 2012 Organizing Committee. https://aclanthology.org/C12-1160/.

Underwood, Ted. 2019. *Distant Horizons: Digital Evidence and Literary Change*. Chicago, IL: University of Chicago Press.

Vliegenthart, Rens, and Stefaan Walgrave. 2008. "The Contingency of Intermedia Agenda Setting: A Longitudinal Study in Belgium." *Journalism & Mass Communication Quarterly* 85, no. 4: 860–77. https://doi.org/10.1177/107769900808500409.

Wang, Liang, Nan Yan, Xiaolong Huang, Linjun Yang, Rangan Majumder, and Furu Wei. 2024. "Multilingual E5 Text Embeddings: A Technical Report." Technical report.

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. 2016. "The FAIR Guiding Principles for Scientific Data Management and Stewardship." *Scientific Data* 3: 160018. https://doi.org/10.1038/sdata.2016.18.

Yablokov, Ilya. 2015. "Conspiracy Theories as a Russian Public Diplomacy Tool: The Case of Russia Today (RT)." *Politics* 35, nos. 3–4: 301–15. https://doi.org/10.1111/1467-9256.12097.

https://doi.org/10.1017/chr.2026.10031 Published online by Cambridge University Press

Topic modeling definitions

Uniform manifold approximation and projection (UMAP): It is a dimensionality reduction technique that is particularly effective for visualizing high-dimensional data in a lower-dimensional space, such as 2D.

Class TFIDF: Technique useful when different classes may have distinct word distributions.

Maximal marginal relevance (MMR): Technique used to balance the relevance and diversity of selected items. Considered a "Representation Model."

KeyBERT: A BERT-specific approach to generating context-relevant keywords to clusters. Also considered a "Representation Model."

Macro-narratives

Diplomacy: Under this Macro-Narrative, we included Meso-Narratives that included topics, such as bilateral relations and cooperation, international organizations dialogue and agreements, treaties, nuclear diplomacy, peace negotiations, and diplomatic appointments, among others.

Politics and society: Under this narrative, we included topics related to domestic politics and society from any country. For example, governance, elections, government affairs and leadership, and controversies with politicians, among others.

Economy: Under this narrative, we included any development in national or international economy. Topics included trade agreements, economic relations, oil industry, economic cooperation, tariffs or trade restrictions, economic crisis or development, and budgets and spending, among others.

Health: Under this narrative, topics included pandemics, public health crisis, vaccines, COVID-19 developments, quarantines, or any other health-related topics.

History: Under this narrative, topics included historical events, past wars or conflicts, or historical figures.

Security and conflict: Under this narrative, topics included any military or armed conflict, wars, terrorism, political conflict, and refugees and migration, among others.

Crime: Under this narrative, topics included anything related to cartels or trafficking, organized crime, criminal organizations, fraud and corruption, and money laundering, among others.

Information: Under this narrative, topics included influence operations, disinformation in social media, WikiLeaks, and espionage, among others.

Technology, science, and innovation: Under this narrative, topics included any developments in science and technology, including new weapons systems or new military equipment and technology announcements.

Selected sources

Sputnik Mundo
RT en Espanol
Embassy of Russia in Colombia
Russian Ministry of Foreign Affairs
El Espectador
El Nuevo Siglo
El Universal
Ahi Les Va!
Colombian Presidential Agency for International Cooperation
Colombian Ministry of Foreign Affairs

Cluster label candidates prompt example

system_prompt = "You are a helpful, respectful and honest assistant for labeling topics."

example_prompt = """ I have a topic that contains the following documents: - Traditional diets in most cultures were primarily plant-based with a little meat on top, but with the rise of industrial style meat production and factory farming, meat has become a staple food. - Meat, but especially beef, is the word food in terms of emissions. - Eating meat doesn't make you a bad person, not eating meat doesn't make you a good one.

The topic is described by the following keywords: 'meat, beef, eat, eating, emissions, steak, food, health, processed, chicken'.

Based on the information about the topic above, please create a short label of this topic. Make sure you to only return the label and nothing more. [/INST] Environmental impacts of eating meat """

user_prompt = f"""I have a topic that contains the following documents: document The topic is described by the following keywords: k Based on the information about the topic above, please create 3 short labels of this topic. Make sure you to only return the label and nothing more."""

https://doi.org/10.1017/chr.2026.10031 Published online by Cambridge University Press