Migration Narratives and Media Framing: An Input–Processing–Response Perspective of Media–Audience Associations

Giulia Contu[a], Carla Massidda[a], Francesco Mola[a], Marco Ortu[a], Silvio Tunis[a]

[a] *University of Cagliari, Cagliari, Italy*

Abstract

This study examines how media narratives on immigration in Italy are associated with public discourse in social media comments using an *input–processing–response* (IPR) framework as an observational analytical design. In this setting, media outlet posts are treated as the *inputs*, comment threads as the observable *responses*, and audience cognition and interaction as an interpretive *processing* layer that is not directly measured. We analyze 708 Facebook posts and 166,760 associated comments published by 10 Italian media outlets. Methodologically, we combine Structural Topic Modeling, *Fightin' Words*, regularized association analysis, and sentiment/emotion indicators to examine how post-side narratives are linked to structured response patterns in comment threads. We find systematic variation in outlet framing, ranging from humanitarian to security emphases, while comment threads exhibit greater thematic dispersion and stronger lexical divergence than outlet posts. Regularized models further show that specific post topics, including humanitarian and migration-management themes, are associated with distinct comment-topic mixtures and shifts in lexical salience. Sentiment and hate-related indicators also vary across outlets and topics, with more negative and hostile aggregate response patterns observed around humanitarian and migration-management themes; these indicators are interpreted at the aggregate level only. Taken together, the findings show how media narratives

Email addresses: `giulia.contu@unica.it` (Giulia Contu), `massidda@unica.it` (Carla Massidda), `mola@unica.it` (Francesco Mola), `marco.ortu@unica.it` (Marco Ortu), `silvio.tunis@unica.it` (Silvio Tunis)

are associated with differentiated audience response structures within an observational IPR framework, with implications for media regulation, platform governance, and media literacy.

Keywords: Structural Topic Modeling, Sentiment and Emotion Analysis, Hate Speech, Media Framing, Social Media, Immigration

1. Introduction

Immigration is one of the most salient and polarizing policy domains in contemporary democracies. Public reactions to migration are shaped by a mix of long-run structural factors (economic growth, welfare regimes, labor demand) and short-run shocks (economic cycles, crises, salient events), with downstream consequences for electoral behavior and for the design and implementation of integration policies (Grigorieff et al., 2020). In this context, media outlets play a dual role. First, as *agenda setters* and *framers*, they influence which aspects of migration become salient and how they are interpreted (McCombs and Shaw, 1972; Iyengar and Kinder, 2024; Iyengar, 1994). Second, in the hybrid media environment, audience responses are immediately observable in online comment sections, where selection into outlets, motivated reasoning, and perceptions of bias can drive divergence from media outlets narratives (Stroud, 2011; Vallone et al., 1985; Lord et al., 1979).

Building on this, we use an input–processing–response (IPR) schema as an observational analytical framework that connects classic communication/persuasion 'input–output' models (Hovland et al., 1953) with Stimulus–Organism–Response (S-O-R) approaches (Robert and John, 1982; Mehrabian and Russell, 1974). Within this perspective, we examine how media narratives on migration are linked to audience reactions on social media, while treating *processing* as an interpretive layer between media exposure and observable response patterns rather than as a directly measured mechanism. Accordingly, our design treats media outlet posts as information inputs and user comments as proximal audience responses. The added value of the IPR perspective in this study lies not in causal identification, but in providing a single observational architecture for jointly analyzing media-side narratives, response-side structures, and their patterned associations. Without this perspective, the empirical strategy would reduce to parallel analyses of posts and comments, with no principled basis for aligning the unit of exposure with the unit of observed response, for treating post–comment links as a substantive object of analysis, or for interpreting heterogeneous empirical results within a common design. The framework also connects our empirical measures to established mechanisms that guide interpretation: agenda setting, framing, and priming on the media side (McCombs and Shaw, 1972; Iyengar and Kinder, 2024; Iyengar, 1994); selective exposure and hostile-media perceptions on the audience side (Stroud, 2011; Vallone et al., 1985); and affective dynamics linked to negativity and perceived threat (Soroka, 2014; Baumeister et al., 2001; Stephan and Stephan, 2000). We also consider meso-level comment-thread dynamics, where opinion leaders and engagement cues may amplify divisive content (Katz et al., 2017; Shaw et al., 1999; Kasperson et al., 1988; Cohen, 1972). These perspectives motivate our measurement choices—topics and frames in posts; topics, sentiment, and hate indicators in comments; and thread-level dispersion—as complementary descriptors of how media inputs relate to structured response patterns within associated discussion threads.

Empirically, we focus on the Italian news ecosystem on Facebook, where media pages serve as both a distribution channel for media outlet content and a venue for public discussion. This setting allows us to compare topical emphases in posts with those emerging in user comments and to study how media-side framing relates to differentiated thread-level response structures, including thematic dispersion and variation in emotional tone. Our approach complements recent computational studies of migration discourse by jointly examining media supply and audience response within the same threads (Drouhot et al., 2023) under a common observational design.

We address the following research questions:

- **RQ1:** Do specific media outlets emphasize particular terms or topics within the immigration debate more than others?
- **RQ2:** Are media topics echoed in public discourse, or do user comments introduce additional perspectives?
- **RQ3:** To what extent do media outlet identity and post-topic composition associate with the distribution of comment topics?
- **RQ4:** How do outlets and frames relate to variations in the emotional tone of public discussions?

The paper makes three main contributions. First, it provides an integrated, thread-level design that links outlet posts and user comments within the same social media platform discussions, Facebook in our study, across multiple Italian media outlets, allowing us to quantify how topical emphases in media inputs are echoed, transformed, or dispersed in audience responses. Second, it translates the Input–Processing–Response framework into an empirical observational design by combining Structural Topic Modeling with regularized association analysis to estimate how outlet identity and post-topic composition relate to comment-topic distributions, while integrating topic structure, lexical divergence, dispersion, and affective indicators into a coherent input–response framework rather than treating them as separate descriptive analyses. Third, it combines topic structure with sentiment and hate-speech indicators to relate specific frames and topics to variation in the emotional tone and hostility of public discourse, yielding a transparent, reproducible description of migration talk in a major European context that informs debates on media practices, platform governance, and media literacy.

The remainder of the paper is organized as follows. Section 2 reviews prior research on media coverage of migration and public discourse on-line. Section 3 presents the data and Section 4 the methods used to analyze the data. Section 5 reports the findings and discusses implications and limitations. Finally, Section 6 concludes.

2. Related Works and Research Questions

2.1. Related works

Research on migration discourse increasingly relies on digital traces from social media to study public attention, attitudes, and reactions to real-world events (Rowe et al., 2022; Righi, 2019; Merrill and Åkerlund, 2018). Within this literature, two methodological streams are central. First, *sentiment and hate/toxicity* measurement pipelines aim to quantify evaluative tone toward migrants and policies. Studies document how threat-laden narratives correlate with hostile language online and how sentiment varies across outlets and audiences (Flores, 2017; Burnap and Williams, 2015; Fortuna and Nunes, 2018; Waseem and Hovy, 2016; Davidson et al., 2017; Liu, 2012; Wilson et al., 2005). Second, *topic modeling* and related text clustering are used to map issue agendas and frames across media and users, including in migration contexts (Helbling, 2014; Viola and Verheul, 2020; Calderón et al., 2020; Gualda and Rebollo, 2016; Kelling and Monroe, 2023). Together, these approaches show that migration discourse is heterogeneous and event-sensitive, with marked shifts around salient incidents and policy changes (Rowe et al., 2022). A complementary strand examines how media frames relate to public discussion. Comparative work shows that security and economic frames tend to elicit more divisive engagement than humanitarian or integration frames (Helbling, 2014; Bartlett and Norrie, 2015). More broadly, political communication research documents how online interactions can sort users into like-minded clusters and intensify affective polarization (Conover et al., 2011; Burnap and Williams, 2016). Recent contributions from Italy emphasize the interplay between media narratives and audience perceptions, including the role of identity, misinformation, and perceived threat (Miccoli et al., 2023; Menshikova and van Tubergen, 2022; Hurtado Bodell et al., 2024). Our study aligns with this literature by jointly describing (i) the topical focus of media outlet posts and (ii) the topics and tone emerging in the corresponding comment threads, treating the latter as observable audience responses in the same conversational space (Drouhot et al., 2023; Garland et al., 2022; Pierri, 2024). Methodologically, we combine standard tools for lexical comparison and topic extraction (Monroe et al., 2008) with sentiment and hate indicators widely used in computational social science (Fortuna and Nunes, 2018; Liu, 2012).

2.2. Theoretical underpinnings and IPR framing

Immigration-related media coverage can be conceived as a sequence that begins with what media outlet posts make salient and how they frame the issue (McCombs and Shaw, 1972; Entman, 1993; Scheufele and Tewksbury, 2007). Audience members are not passive recipients of these cues: they differentially attend to, select, and interpret content in line with prior attitudes and preferences, shaping actual exposure (selective exposure) (Stroud, 2011). Once encountered, messages may activate evaluative considerations (priming) (Iyengar, 1994; Iyengar and Kinder, 2010) and elicit asymmetric reactions to negative or threat-related information (negativity bias) (Baumeister et al., 2001; Soroka, 2014). In the immigration domain, threat-based interpretations and episodes of heightened alarm can become particularly relevant, as suggested by integrated threat theory and moral panic accounts (Stephan and Stephan, 2000; Cohen, 1972). Finally, media influence is plausibly filtered and transformed through social interaction—for instance via opinion leaders and interpersonal networks (two-step flow) (Katz et al., 2017)—and through endogenous thread-level dynamics that shape what becomes visible and contested in discussion spaces. The downstream footprint of these processes is visible in comment-thread language through observable patterns:

what topics are foregrounded, how posts are rephrased or contested, and which affective tones emerge.

Media-effects and social-psychological theories thus provide complementary lenses for interpreting how immigration-related media content co-occurs with patterns of audience discussion on social media. In this study, the empirical goal is to characterize post/thread-level associations between media outlet posts and the ensuing comment-thread language. Accordingly, we focus on what is directly observable in our data: (i) the textual content of media posts (*inputs*) and (ii) the distribution and tone of comments within the corresponding threads (*responses*). Building on this literature, we use the *input–processing–response* (IPR) schema in Figure 1 as an organizing framework that separates operationalized constructs (inputs and observed comment-thread responses) from theorized constructs used to guide interpretation of the empirical input–response associations.

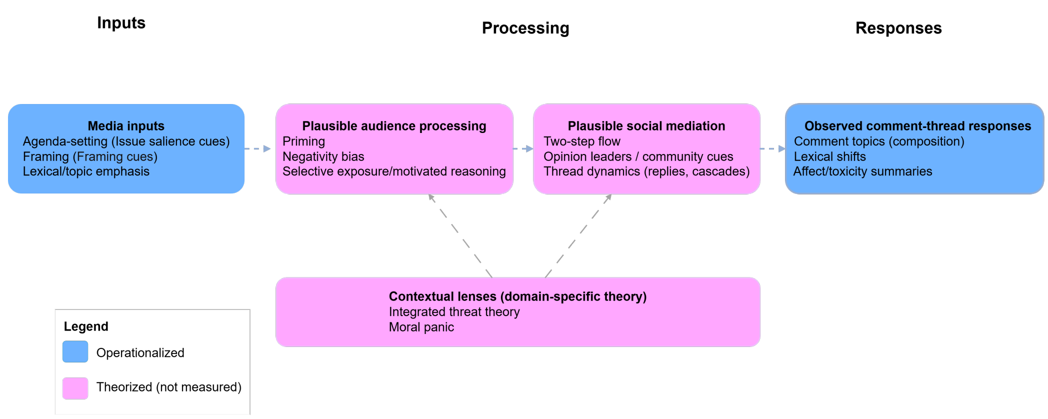

Figure 1: IPR overview linking theoretical lenses to observable constructs. Blue elements denote constructs **operationalized** in this study (media outlet posts as inputs; comment-thread measures as responses). Pink elements denote **theorized** cognitive and social mechanisms and domain-specific contextual lenses used for interpretation only. Dashed arrows indicate plausible links discussed in prior literature and represent interpretive paths rather than identified effects.

Figure 1 summarizes the interpretive structure of the empirical design, while empirical inference remains focused on post/thread-level associations.

2.2.1. Interpretive lenses (conceptual constructs)

Media inputs and media-effects perspectives. Agenda-setting, framing, and priming are central media-effects perspectives that describe how media content can shape salience, interpretation, and the accessibility of evaluative considerations (McCombs and Shaw, 1972; Entman, 1993; Scheufele and Tewksbury, 2007; Iyengar and Kinder, 2010; Iyengar, 1994). We do not treat these theories as a set of sequential stages. Rather, they motivate attention to observable features of media outlet posts—including issue salience cues, framing cues, and lexical/topic emphasis—and provide a vocabulary for interpreting patterns of post–comment alignment and lexical uptake.

Plausible audience processing (not measured). Selective exposure and motivated reasoning emphasize that exposure and interpretation depend on prior attitudes and preferences, including phenomena such as hostile-media perceptions and counter-arguing (Stroud, 2011; Vallone et al., 1985; Lord et al., 1979). Negativity bias and affect heuristics further suggest asymmetric reactions to negative, threatening, or alarmist cues (Soroka, 2014; Baumeister et al., 2001). We use these constructs to interpret heterogeneity in observed input–response associations across outlets, topics, and time.

Plausible social mediation (not measured). Beyond individual processing, two-step flow and related diffusion perspectives highlight that media influence can be filtered and reinterpreted through opinion leaders and interpersonal networks (Katz et al., 2017). In platform settings, endogenous thread dynamics (replies, cascades, local reinforcement) may further shape which cues become prominent in the discussion. These perspectives help contextualize why comment threads may amplify, contest, or reframe post content.

Domain-specific contextual lenses. Immigration is a substantively distinctive domain in which threat-based interpretations and episodic amplification are plausible. Integrated threat theory distinguishes realistic and symbolic threats as drivers of outgroup hostility (Stephan and Stephan, 2000), while moral panic and related amplification frameworks describe episodic surges in attention and alarm around salient incidents (Cohen, 1972; Kasperson et al., 1988). In Figure 1, these aspects are treated as *contextual lenses* rather than stages of processing: they inform interpretation of patterns (e.g., threat-laden lexical shifts or event-driven concentration) without being operationalized as measured components.

Observed comment-thread responses (empirical focus). The outcomes observed in our setting are properties of the comment threads associated with each post: topical composition (issue attention), lexical shifts relative to the post, affect/toxicity summaries, engagement signals, and concentration/dispersion diagnostics. We interpret results as changes in *thread composition and tone*, not as individual-level attitudinal effects.

2.2.2. Operationalization (measured constructs) and empirical mapping

Inputs (measured). Inputs are measured from media outlet posts via post-level topic mixtures and lexical/emphasis cues that operationalize aspects of salience and framing. Agenda-setting and framing are invoked here as media-effects perspectives motivating the measurement of salience and framing cues in posts; they are not treated as tested stages.

Processing and mediation (not measured). We do not estimate processing or social mediation mechanisms directly. Empirically, we model conditional associations between input features (including outlet identity and post-topic composition) and response-side thread summaries using a regularized association analysis (elastic-net). Coefficients are interpreted descriptively and in light of the interpretive lenses outlined above.

Responses (measured). Response measures are computed from comments and aggregated at the thread level to align the unit of analysis with the unit of exposure (the post). They include comment-topic shares, affect/toxicity indices, lexical divergence from posts, dispersion diagnostics, and engagement metrics.

Dispersion and divergence diagnostics. We use response-side dispersion to characterize the internal concentration versus spread of comment-thread content, and lexical divergence diagnostics to quantify the extent to which comment threads depart from the initiating post. Operationally, these diagnostics include topical entropy and Jensen–Shannon divergence between post and comment lexical distributions. We discuss higher dispersion or divergence as consistent with user agency and cross-cutting exposure documented in prior work, without attributing either pattern to a single mechanism.

Input–response mapping. Table 1 summarizes how observable post-level inputs are linked to response-side indicators, stated as *interpretive expectations* grounded in prior literature.

Empirical grounding The above interpretive expectations align with prior evidence that media emphasis can correlate with issue salience and affective reactions on social platforms, while user agency and cross-cutting exposure

Input	**Interpretive mechanism (theorized)**	**Hypothesized association patterns**
Issue topics (agenda)	Agenda-setting; availability/priming	Higher topic salience in posts is expected to be associated with stronger post–comment topic alignment and higher issue-attention share in comments.
Frames (problem-attribution-solution; valence)	Framing and appraisal; motivated reasoning	Frame emphasis is expected to be associated with systematic shifts in affective tone and distinctive lexical deltas; negative/attribution frames may co-occur with higher toxicity/hate indicators.
Threat cues (realistic/symbolic)	Integrated threat theory	Higher threat cue prevalence is expected to be associated with increased outgroup/hostile language, higher toxicity/hate indicators, and lexical shifts toward threat-related terms.
Negativity / intensity	Negativity bias; affect heuristic	More intense or negative post emphasis is expected to be associated with higher engagement and stronger affective responses than positive emphasis.
Outlet–audience incongruence	Selective exposure; hostile media	Outlet–audience mismatch is expected to be associated with weaker topical alignment and more adversarial lexical shifts.

Table 1: Input–processing–response mapping and hypothesized association patterns. Mechanisms represent interpretive lenses from prior literature and are not directly observed or estimated in this study.

can attenuate direct transmission (Kim et al., 2025; Zhang and Lian, 2025; Yarchi et al., 2024).

3. Data collection

The dataset analyzed in this study comprises Facebook posts and comments about immigration in Italy collected from ten major Italian media outlets, from February 2022 to April 2024. We focus on Facebook because it combines high reach in Italy with public, threaded discussions on outlet pages, enabling measurement of audience responses at the thread level. Regulatory audience statistics document substantial monthly usage of social networking services in Italy (unique users and time spent), underscoring Facebook's persistent penetration in the media diet (AGCOM, 2024b,a). In parallel, the Reuters *Digital News Report* shows that social platforms remain important gateways for news, including in Italy, despite recent platform resets and compositional shifts in news pathways (Reuters, 2024, 2025). Crucially for this study, Facebook is a central arena for political and migration communication in the Italian context: leaders and parties systematically use Facebook to discuss immigration (Prislei et al., 2022; Mazzoleni and Bracciale, 2018), and migration-related political messaging and targeting have been extensively analyzed using the Facebook Ads Library in Italy (Capozzi et al., 2021). The page architecture exposes full, nested comment threads on public outlet posts, allowing us to link *inputs* (outlet posts) to observable *responses* (comment topics, stance, and toxicity) within our input–processing–response design. The selection of media outlets was carefully designed to represent diverse perspectives on immigration, varied geographical focus, and different audience demographics. *Corriere della Sera*, *Libero*, *Il Fatto Quotidiano*, *Il Giornale* and *La Repubblica* were included as they represent the media outlets with the highest nationwide circulation. To capture regional perspectives while maintaining national relevance, *Il Resto del Carlino* and *Il Mattino* were selected, as these outlets provide national coverage with particular attention to their respective geographical areas. The study also includes digital-only media outlets through *Il Post* and *Fanpage.it*, with the latter being the most widely used online media outlet in Italy as of 2024, according to the Reuters Institute Digital News Report. Additionally, *Avvenire*, with its Catholic-inspired editorial approach, was included to represent a specific demographic with distinctive values on humanitarian issues. The data collection process employed a semi-automated web scraping methodology developed specifically for this research. This approach began with the identification of the Facebook page for each selected media outlet. Subsequently, a search was conducted within these pages for posts containing relevant immigration-related keywords, including "migranti," "immigrazione," "sbarchi," "ong," "immigrati," "lampedusa," "extracomunitari," "mediterraneo," and "scafisti". To ensure data quality and relevance, a manual filtering stage, reviewing each post to confirm it specifically addressed immigration issues rather than merely mentioning the keywords in passing or in unrelated contexts. The initial number of posts obtained in this phase is 923. Following this verification, the URLs of relevant posts were systematically extracted. For each post, the comment sections were expanded to reveal all comments, including nested replies that are typically hidden in the Facebook interface. Finally, the text content of all comments was extracted and stored for analysis. The final dataset comprises 708 posts across all ten media outlets, with 166,760 associated comments. For topic modeling only, we subsequently removed posts with fewer than five tokens (analysis-stage filtering), resulting in **598** posts.

media outlet	Posts	Comments
Avvenire	64	2,276
Corriere della Sera	48	19,622
Fanpage.it	48	23,397
Il Fatto Quotidiano	109	25,522
Il Giornale	108	22,382
Il Mattino	62	9,744
Il Post	79	6,791
Il Resto del Carlino	41	2,988
La Repubblica	57	33,241
Libero	92	20,797
TOTAL	**708**	166,760

Table 2: Distribution of posts and comments across media outlets

The distribution of posts and comments, as shown in Table 2, varies significantly across outlets, reflecting differences in audience engagement and posting frequency. *Il Giornale* contributed the largest number of posts (108), followed by *Il Fatto Quotidiano* (109) and Libero (92). In contrast, *Il Resto del Carlino* had the smallest representation with 41 posts, while *Corriere* *della Sera* and *Fanpage.it* each contributed 48 posts. The distribution of comments shows a greater variation. *La Repubblica* garnered the highest engagement with 33,241 comments despite having only 57 posts. This suggests an average of approximately 583 comments per post, significantly higher than the dataset average. *Il Fatto Quotidiano* and *Fanpage.it* also received substantial engagement with 25,522 and 23,397 comments, respectively. At the lower end of the engagement spectrum, *Avvenire* received only 2,276 comments across its 64 posts, and *Il Resto del Carlino* gathered 2,988 comments on its 41 posts. This variation in comment volume likely reflects differences in audience size, demographic composition, and the engagement strategies employed by each media outlet.

4. Methodology

This section describes the empirical design used to address the research questions (RQs) and situates the corresponding constructs within the input–processing–response (IPR) overview in Section 2.2. In our data, media posts provide the observable *inputs*, while comment-thread language (and thread-level interaction summaries) provide observable *responses*. Intermediate *processing* and *social mediation* constructs are not measured and are not modeled as mediators; they appear in the IPR overview as interpretive lenses (dashed links) used to organize the descriptive linkage between input-side narratives and observable response-side thread structures. Accordingly, we estimate separate topic models for outlet posts (inputs) and for comments (responses), and quantify post–comment relationships via a regularized association analysis.

4.1. Inputs: Lexical Salience by Outlet (RQ1)

RQ1 asks whether specific media outlets disproportionately emphasize particular terms or themes. We operationalize this input-side question via a lexical salience analysis on outlet posts using the log-odds ratio with an informative Dirichlet prior (*Fightin' Words*) (Monroe et al., 2008). The design is one-versus-rest at the outlet level on the corpus of *posts only*: for each outlet, we compare the outlet's post-term counts against pooled counts from all remaining outlets after a shared preprocessing pipeline. The estimator yields outlet-specific log-odds shifts and associated z-scores that quantify which terms are disproportionately used by a given outlet relative to the rest, providing an observational measure of issue- and framing-related lexical emphasis in the media inputs. Full definitions, priors, variance approximation, and display conventions are reported in Appendix A.4.

To ensure interpretability as an input-side construct, no information from comment threads enters this procedure. We trim rare terms, standardize preprocessing across outlets, and report the top-m positive and negative terms per outlet by z-score together with representative examples. Empirical results addressing RQ1 are presented in Section 5.1.

4.2. Inputs versus Responses: Post Topics and Comment Topics (RQ2)

RQ2 asks whether media topics are echoed in public discourse or whether user comments introduce additional perspectives. We address this by estimating *separate* Structural Topic Models (STM) for outlet posts (inputs) and for comment threads (responses), and then comparing the resulting thematic inventories and prevalence summaries. Estimation for both corpora follows the logistic-normal STM; implementation details are reported in Appendix A.2. The number of topics in each model is selected by balancing held-out predictive performance, semantic coherence, and exclusivity, with diagnostics defined in Appendix A.2.1 and reported in Figures 3–5. For the comments STM, the prevalence specification includes outlet indicators and the parent post identifier (`id_post`) to account for systematic heterogeneity in topic prevalence across outlets and threads.

The comparative design proceeds in two steps. First, we construct topic inventories for posts and comments (top terms, representative documents, and human-readable labels), using the same preprocessing and selection criteria on each side (Appendix A.2.3). Second, we summarize topic prevalence at the corpus and outlet-stratified levels for each model and, where substantive correspondences between post topics and comment topics can be supported by side-by-side inspection of top terms and exemplar texts, we report these as qualitative alignments. These contrasts are used to characterize echoing versus expansion as observable patterns of response-side differentiation relative to media inputs and do not rely on cross-model distance metrics.

Auxiliary shared-space lexical divergence (thread-level). In addition to prevalence comparisons, we quantify the extent to which comment threads lexically track the initiating post versus introduce additional lexical material. For each post p, we build normalized term-frequency distributions P_p (post text) and

Q_p (aggregated comments) over the post-STM vocabulary V_{post} by representing both texts in the shared lexical support induced by V_{post}. We then compute the Jensen–Shannon distance $D_{\text{JS}}(P_p, Q_p)$. This yields a symmetric divergence measure in a common coordinate system (the post vocabulary) without requiring identification or matching of latent topic axes across separate models. This auxiliary diagnostic is defined only on the shared support induced by the post vocabulary and does not alter the main analysis, in which the full comment corpus is modeled separately. Token coverage diagnostics and full results are reported in Appendix A.7.

4.3. Input–Response Association via Elastic-Net (RQ3)

To quantify how inputs relate to responses, we model, for each comment topic t, the thread-level topic share as a function of post-level predictors comprising outlet indicators and the post's topic composition. The penalized linear specification employs an elastic-net penalty with $\alpha = 0.5$ and λ chosen by 10-fold cross-validation. Estimation, predictor standardization, penalty-path selection, and uncertainty quantification via post-level bootstrap and cluster-robust summaries are provided in Appendix A.3. This formulation directly addresses RQ3 by estimating the extent to which outlet identity and post-topic composition are associated with the topical allocation observed in comment threads.

We interpret the resulting associations descriptively and discuss them in relation to the theoretical lenses outlined in Section 2.2 (e.g., agenda-setting, selective exposure, threat perception), without claiming identification of the corresponding mechanisms. Within the IPR schema, "processing" is therefore not operationalized as a measured mechanism, but treated as an interpretive layer linking input features to structured response-side heterogeneity in post–comment associations.

Unit of analysis and aggregation. Although comments are nested within posts, our estimand is the *thread-level* audience response induced by a given media post, not individual-level commenting behaviour. Accordingly, we summarize each comment thread into post-level outcomes (e.g., comment-topic shares and affective/hostility indices) and treat the post–thread pair as the core empirical unit in downstream models. This choice matches the conceptual framing in Figure 1, where the empirical layer refers to *observed comment-thread responses.* This alignment follows directly from the IPR design, which treats the post as the unit of exposure and the associated comment thread as the observable response, thereby making the post–thread pair, rather than posts and comments in isolation, the core empirical unit of the study. This aggregation mitigates pseudo-replication under within-thread dependence (replies and conversational cascades) and prevents a small number of high-volume threads from mechanically dominating estimation.

4.4. Inputs versus Responses: Emotional Tone and Hostility Indices (RQ4)

RQ4 investigates how outlets and framing-related cues in posts (approximated by topic composition and lexical emphasis) relate to variation in the emotional tone of public discussions. We treat the response as a set of affective and hostility summaries computed on comments and aggregated at the thread level, and then compare these outcomes across outlets and outlet×topic cells to relate input-side variation (outlet identity and post-level framing cues) to response-side tone. Index definitions are reported in Appendix A.5 along with model evaluation in Appendix A.5.1; empirical results are presented in Section 5.4.

At the comment level, we obtain sentiment polarity and activation from a validated classifier and detect hate/hostility with calibrated severity thresholds (Appendix A.5). These predictions are aggregated into standardized outcomes: *Polarity Index by outlet* (PI_m), *Intensity Index by outlet* (II_m), *Hate Intensity Index by outlet* (HII_m), *Hate Polarization Index by outlet* (HPI_m), and *Hate Polarization Index by outlet-topic* (HPI_{mp}). Each index is computed from thread-level aggregates and standardized against a global baseline as defined in Appendix A.5.

To relate inputs to these responses, we summarize PI_m, II_m, HII_m, and HPI_m at the outlet level and HPI_{mp} across the outlet×topic grid. This design isolates outlet identity and post-level content variation as input-side factors while avoiding information flow from responses into the construction of inputs.

4.5. Validation and Reproducibility

The analytical pipeline proceeds from corpus preprocessing to topic estimation for posts and comments, continues with post–comment association modeling, and concludes with validation and robustness checks. Text normalization, tokenization, lemmatization, stopwording, bigram construction, and trimming rules are described in Appendix A.1.

Component	Setting	Rationale
STM (posts; inputs)	K=5; logistic-normal prior; spectral init; 250 iterations	Selected by coherence, exclusivity, and held-out performance; yields interpretable outlet frames (Figs. 3–5; Appendix A.2.1).
STM (comments; responses)	K chosen via identical diagnostics; same priors; 250 iterations	Harmonized selection supporting downstream association analysis; prevalence covariates include outlet indicators and `id_post` to accommodate thread heterogeneity (Appendix A.2.1).
Text covariates	pre-specified if used	Included only when improving held-out likelihood to limit overfitting (Appendix A.2.1).
Elastic-net (association)	α=0.5; λ via 10-fold CV on a 100-point path	Balances sparsity and grouping; CV minimizes deviance (Appendix A.3).
Unit of analysis	Post-level; thread-level outcomes	Aligns exposure (post) and response (thread) and avoids cross-model topic matching (Appendix A.2.3).
Bootstrap	B=1000 post-level resamples	Percentile CIs and cluster-robust summaries (Appendix A.3).
Seed control	fixed across modules	Exact replicability (Appendix A.6).

Table 3: Main hyperparameters, final settings, and rationales.

Hyperparameters, final settings, and rationales are summarized in Table 3; quantitative diagnostics guiding topic-model selection are presented in Figures 3 and 5 and detailed in Appendix A.2.1. All stochastic procedures (topic initialization, cross-validation, bootstrap resampling) are executed with fixed seeds to ensure exact reproducibility. The computing environment and R/Python scripts are available in a public repository as detailed in Section 10.

5. Results

In this section, we report empirical results organized around the four research questions and the layers of the IPR framework. We first describe input-side variation in outlets' lexical choices (RQ1, Inputs), then compare topic structures in posts and comments to assess topic echo and dispersion (RQ2, Input→Response). Next, we present regularized association models that quantify post–comment topic linkages (RQ3, Input→Response Associations), and finally we examine sentiment and hate-speech indices as affective responses to media inputs (RQ4, Input→Response).

5.1. Lexical Salience by Outlet: Fightin' Words (RQ1)

In this section, we present the results of the Fightin' Words technique described in Section 4 and Appendix A.4. The visualization in Figure 2 suggests that media posts about immigration vary systematically across different media outlets, with each group favouring particular lexical choices. In a "Fightin' Words"-style plot, terms that appear toward the edges of the chart are those strongly associated with a specific outlet's section, whereas more central terms tend to be shared among multiple outlets.

A first distinction concerns posts that emphasize a humanitarian or solidarity-based perspective versus those underscoring crisis or emergency frames. For instance, words such as "salvare," "costiera," "profughi," and "solidarietà" appear clustered together, reflecting a discourse centred on rescue operations and moral obligations. These terms occur disproportionately in outlets that—based on the chart's colour coding—promote or attract discussion around ethics, empathy, and social welfare (e.g., references to Catholic-oriented or progressive media). In contrast, terms like "collasso," "allarme," "emergenza," and "guerra" cluster elsewhere and suggest a more alarmist or conflict-focused framing. Posts from outlets favouring this lexicon appear to stress the disruptive consequences of immigration, often associating

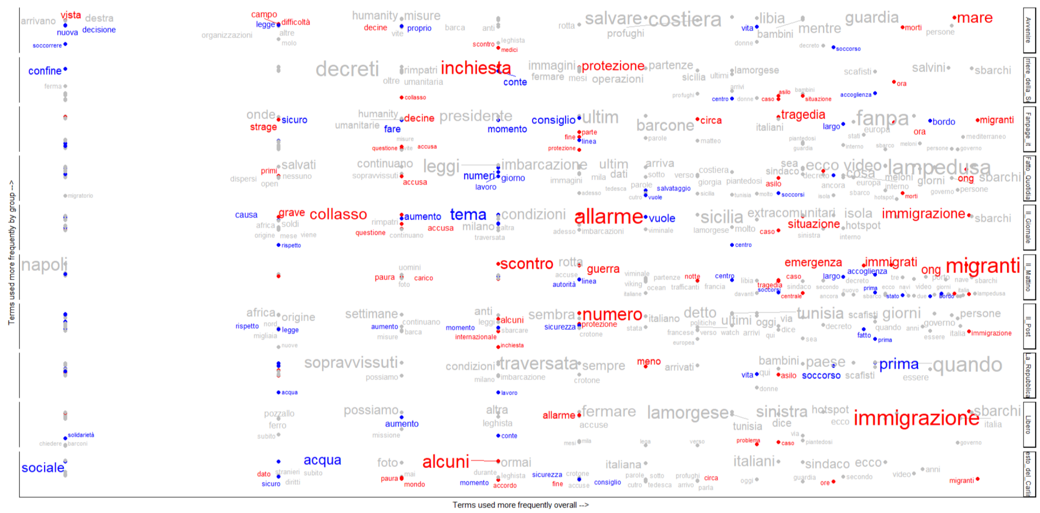

Figure 2: Comparison of user-generated terms across media outlets. The x-axis represents overall frequency of use, and the y-axis differentiates words favoured by specific outlets. Colours indicate sentiment polarity, with red for negative, blue for positive, and grey for neutral. Word size is scaled by the statistical significance of the "Fightin' Words" test Monroe et al. (2008).

arrivals with potential breakdowns of public order and economic capacity. They also use language oriented around "difesa" and "controllo," indicating concern with border security or a perceived need for restrictive legislation. Beyond these two broad orientations, other clusters reveal more localized or thematic focuses. For instance, "napoli," "sociale," or "campo" might indicate outlets (often local media outlets) where posts ground the debate in municipal or regional issues, discussing the impact of immigration on specific cities or highlighting local policies. Relatedly, the presence of words like "africa," "origine," "rottura," and "sbarco" in a separate segment underscores a different thread of conversation: many posts concentrate on geographic origins, routes (e.g., "lampedusa," "tunisia"), and the logistics of arrivals (e.g., "barcone," "rotte," "acqua"). Several outlets also differ in their use of language referencing political actors and legislative measures. Terms like "leggi," "decreti," "presidente," and "conte" appear in a central region, reflecting that legislative discourse—though present in all media outlets—can lean more technical or formal depending on the platform. Meanwhile, mentions of local political figures or more populist buzzwords (e.g., "destra," "sinistra," "sicuro," "scontro") cluster in separate areas, suggesting a sharper partisan framing in some comment sections (often linked to outlets that are editorially aligned with a particular political orientation).

In sum, the analysis conducted so far addresses RQ1 by indicating that certain media outlets place greater emphasis on specific terms or topics within the immigration debate compared to others. The posts reveal substantial lexical variation tied both to editorial orientation (humanitarian vs. alarmist frames), local vs. national focus (municipal concerns vs. large-scale narratives), and degree of policy-focused discourse. Certain outlets speak more frequently in terms of solidarity or humanitarian aid ("soccorso," "profughi"), whereas others are prone to discuss immigration in urgent or conflictual terms ("allarme," "collasso," "scontro"). Still, others place immigration narratives in a local context or emphasize legal and political procedures. These distinctions underscore how the same topic (immigration) elicits qualitatively different language patterns depending on the media environment.

5.2. Topic Structure in Posts and Comments (RQ2)

5.2.1. Media Posts

This part of the study delves into the details of topic analysis of media posts.

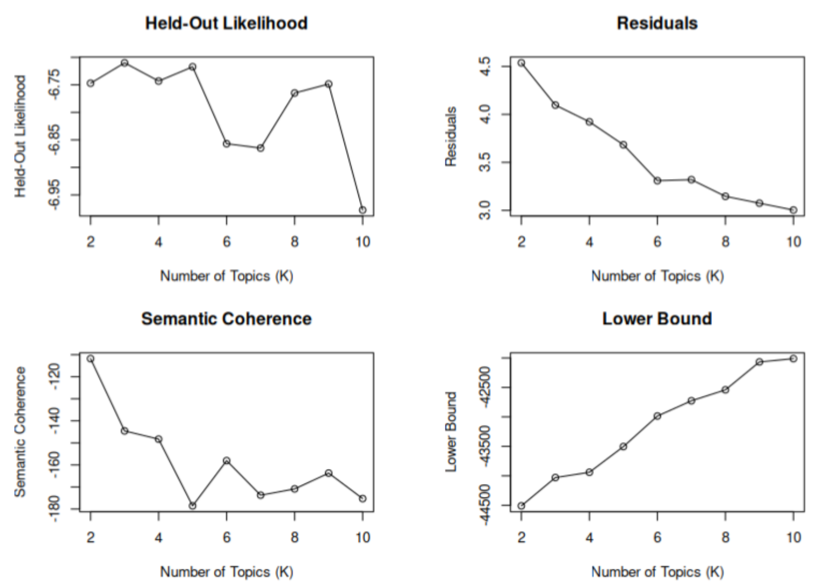

Figure 3: Diagnostic Values by Number of Topics for Media Posts

To determine the optimal number of topics for media posts, we evaluate diagnostic metrics across different values of K (Figure 3). The analysis indicates $K = 5$ as the optimal choice, representing a balance point across multiple diagnostic measures. The residuals plot shows significant improvement until K=5-6, after which the gains become marginal, indicating that five topics adequately capture the main discourse patterns. While semantic coherence is highest at K=2, it stabilizes after K=5, suggesting that additional topics do not substantially improve interpretability. The held-out likelihood displays good performance at K=5 (-6.71) before degrading with higher K values. The lower bound metric also exhibits diminishing returns after K=5, further supporting this choice.

K	Held-out	Residual
4	-6.657	3.816
5	**-6.984**	**3.133**
6	-6.716	3.236

Table 4: Posts topic-number sensitivity ($K \in \{4, 5, 6\}$).

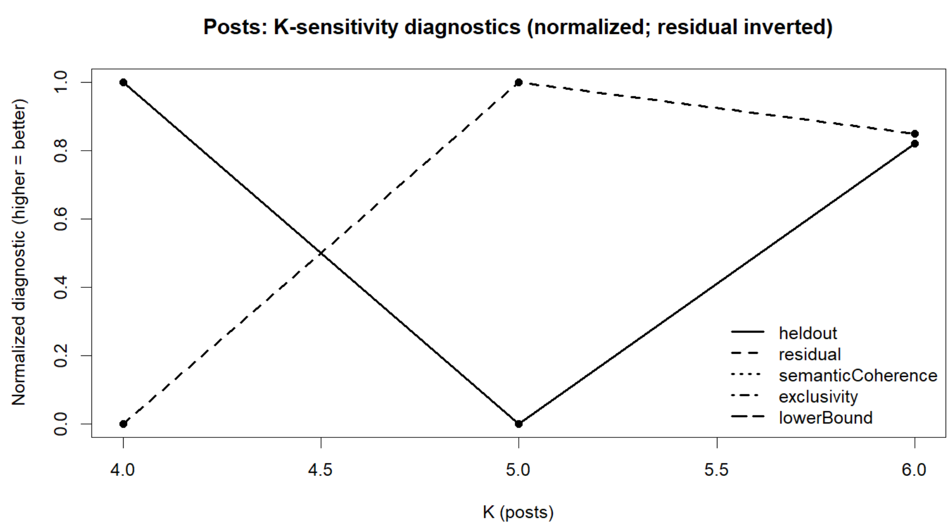

Figure 4: Posts STM diagnostics across $K \in \{4, 5, 6\}$. Metrics are normalized to $[0, 1]$ for comparability; residuals are inverted so higher values indicate better fit.

To assess sensitivity of this choice, we additionally evaluate $K \in \{4, 5, 6\}$ using the same diagnostic suite (Table 4; Figure 4). The results indicate a trade-off across criteria: residual-based fit improves from $K = 4$ to $K = 5$ and does not materially improve at $K = 6$ (Table 4), while the normalized diagnostics show no consistent gain from increasing K beyond 5 (Figure 4). Given the limited size of the post corpus, larger K values risk producing fragmented, low-prevalence topics that are less stable across runs and harder to interpret as outlet-level frames, with limited incremental fit. We therefore adopt $K = 5$ as a parsimonious solution for the smaller post corpus, balancing fit and interpretability and yielding stable, substantively coherent topics for outlet-level framing analysis.

Table 5 presents these topics with their most representative words. It is worth noting that the topics identified in our analysis do not align precisely with the framing categories commonly referenced in the literature on immigration discourse (Kriesi, 2012; Helbling, 2014). Using the five frame categories proposed by Kelling and Monroe (2023) as a point of reference (identity, morality, economics, security, legality), we contend that each of our identified topics intersects with two or more of these framing dimensions.

#	Topic	Most Representative Terms
1	Sea Rescue Operations	ong, mediterraneo, mare, nave, porto, bordo, morti, soccorso
2	Humanitarian Aspects	vita, watch, profughi, immagini, alcuni, avvenire
3	Government Policy	salvini, immigrati, decreto, frances, protezion, scafisti
4	Migration Management	sbarchi, hotspot, meloni, accoglienza, emergenza, isola
5	European Politics	migranti, interno, europa, sindaco, francia, piantedosi

Table 5: Topic Composition and Representative Terms in Media Posts

5.2.2. User Comments

We move now to the selection of the optimal number of topics for user comments. This operation requires balancing multiple diagnostic metrics (Figure 5). While the held-out likelihood shows continuous improvement up to K=30, the residuals metric suggests an optimal point at K=25 (230.26) before degrading at K=30 (271.34). Semantic coherence decreases steadily as K increases, indicating a trade-off between model fit and interpretability. The lower bound shows diminishing returns after K=15, with marginal improvements for higher values. Given these patterns, K=15 emerges as a reasonable choice: it captures substantial improvement in model fit (as shown by the lower bound), maintains better semantic coherence (-190.21) compared to higher K values, and shows good held-out likelihood performance (-8.37). Compared to the results in Kelling and Monroe (2023), our analysis highlights a lower number of topics in readers' discourses (15 instead of 24).

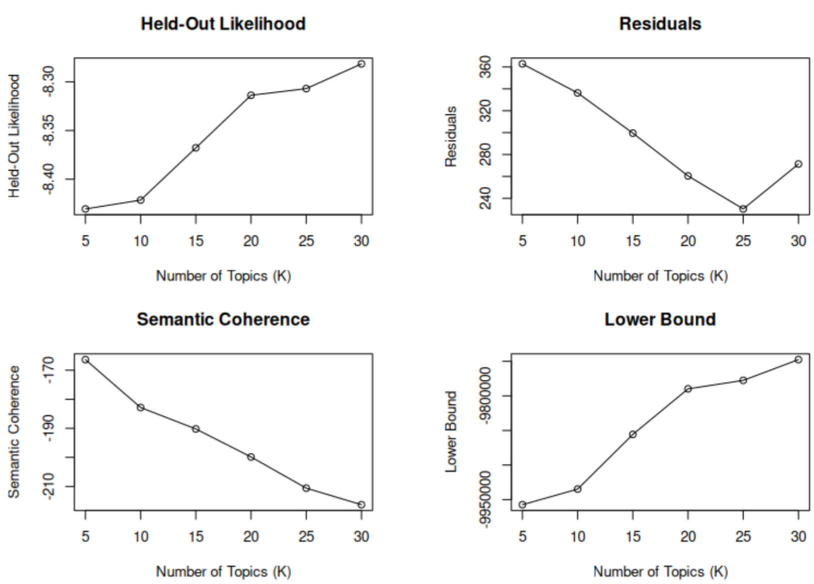

Figure 5: Diagnostic Values by Number of Topics for User Comments

Table 6 shows the topics' most informative keywords. This general picture highlights a more diverse and distributed public discourse compared to media posts. While media coverage focuses on five main topics, public discourse expands into various additional dimensions, including *Social Impact* (Topic

7), *Domestic Issues* (Topic 8), *Political Discourse* (Topic 9), and *Economic Impacts* (Topic 13).

These findings provide a first response to RQ2, showing how user comments introduce novel perspectives into their discourses. The analysis suggests that comment sections often serve as forums for diverse viewpoints that may not align with their host media outlets' editorial stances.

5.2.3. Topic Prevalence between Media Posts and Comments

To further support our response to RQ2, we now focus on the topic prevalence comparison between the content of the media posts and the comments of the users. This analysis aims to highlight the similarities and differences in thematic focus between journalistic narratives and audience reactions, thereby providing a deeper understanding of the dynamics between media discourse and public engagement.

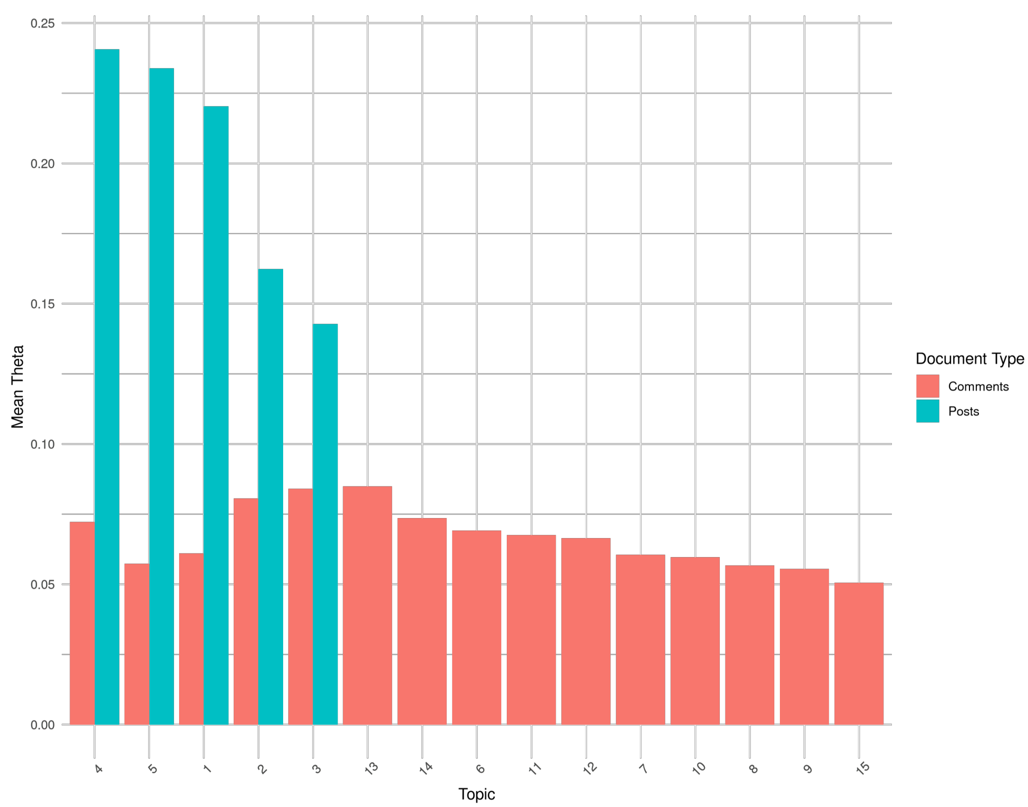

Figure 6: Topic Prevalence Comparison Between Posts and Comments

#	Topic	Most Representative Terms
1	General Discussion	dire, problema, qui, niente, forse, tempo, fine
2	Maritime Operations	ong, mare, porti, nave, navi, francia, germania, sbarcare
3	Political Leadership	governo, salvini, ora, sbarchi, ministro, destra, conte
4	Public Opinion	fatto, bene, mai, proprio, dice, certo, ragione
5	Policy Proposals	fare, cosa, meloni, adesso, blocco, navale, umani
6	Temporal Perspective	così, ancora, anni, già, posto, poco, giusto
7	Social Impact	poi, sempre, prima, gente, legge, parla, povera
8	Domestic Issues	casa, paese, può, invece, visto, detto, normale
9	Political Discourse	solo, quando, allora, fuori, parole, lega, dublino
10	Anti-Immigration Sentiment	migranti, basta, clandestini, italiano, ogni, arrivano
11	Immigration Laws	essere, altri, colpa, vero, stati, immigrazione
12	European Dimension	italia, europa, deve, pure, andare, vai, stare
13	Economic Impact	italiani, soldi, lavoro, africa, immigrati, guerra
14	Left-wing Criticism	sinistra, avere, vita, grazie, vergogna, politica
15	Systemic Issues	senza, persona, stato, vuole, nessuno, mondoS

Table 6: Topic Composition and Representative Terms in User Comments

Figure 6 presents a comparative analysis of topic prevalence between media posts and user comments. The distribution shows a clear structural difference: while in media posts, some of the five topics carry significantly higher prevalence ($\theta \in [0.14, 0.24]$), the distribution of topic prevalence in user comments appears much more balanced ($\theta \in [0.05, 0.08]$). Among media posts, *Migration Management* (Topic 4) leads with the highest expected proportion, followed closely by *European Politics* (Topic 5) and *Sea Rescue Operations* (Topic 1). *Humanitarian Aspects* (Topic 2) and *Government Policy* (Topic 3) show relatively lower proportions but still maintain a substantial presence in the media narrative. In contrast to previous literature, moral frames (topics 1, 2), though highly important, do not emerge as the most dominant in our findings. Instead, security and legality frames (topics 1, 3, 4, 5) are overwhelmingly prominent within media discourse. Conversely, identity and economic frames appear to receive comparatively limited attention. Among user comments, the topics display a more balanced distribution. The rightmost panel of Figure 7 reveals that while *Economic Impact* (Topic 13), *Political Leadership* (Topic 3), and *Maritime Operations* (Topic 2) lead in prevalence, the differences between topic proportions are much smaller compared to media posts.

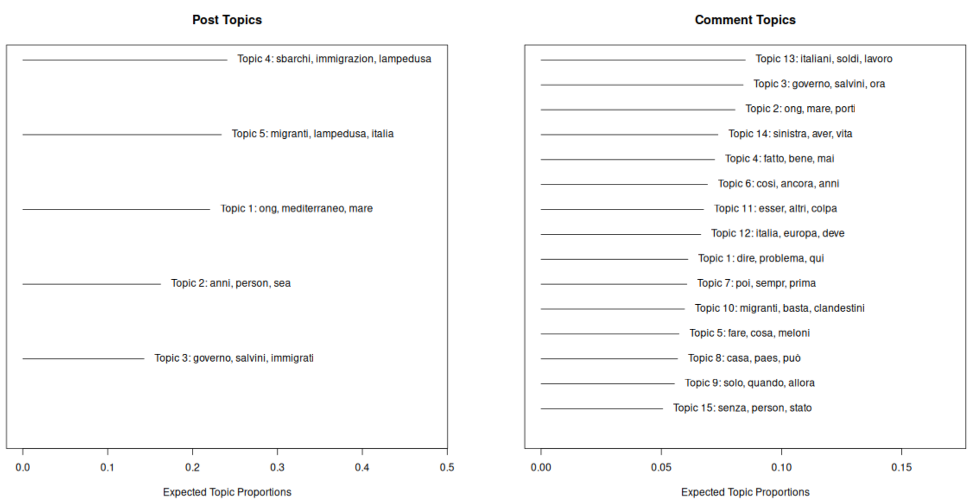

Figure 7: Expected Topic Proportions in Posts and Comments

The expected topic proportions reported in Figure 7 are consistent with patterns in this scenario. While media coverage focuses heavily on operational and policy aspects, public discourse expands into various additional dimensions. The lower prevalence values but broader topic distribution in comments ($\sigma^2_{\text{comments}} < \sigma^2_{\text{posts}}$) confirms that public discourse tends to fragment the immigration debate into more specific subtopics. In other words, while media coverage sets the main themes of discussion, public discourse develops these themes into more nuanced and varied conversations.

This is further evident in Figure 8, which presents the entropy distributions for both media posts and user comments, revealing distinct patterns in topic mixing. The entropy measure $H(\theta)$ quantifies the degree of topic diversity within each document, where higher values indicate more balanced topic mixtures and lower values suggest topic specialization.

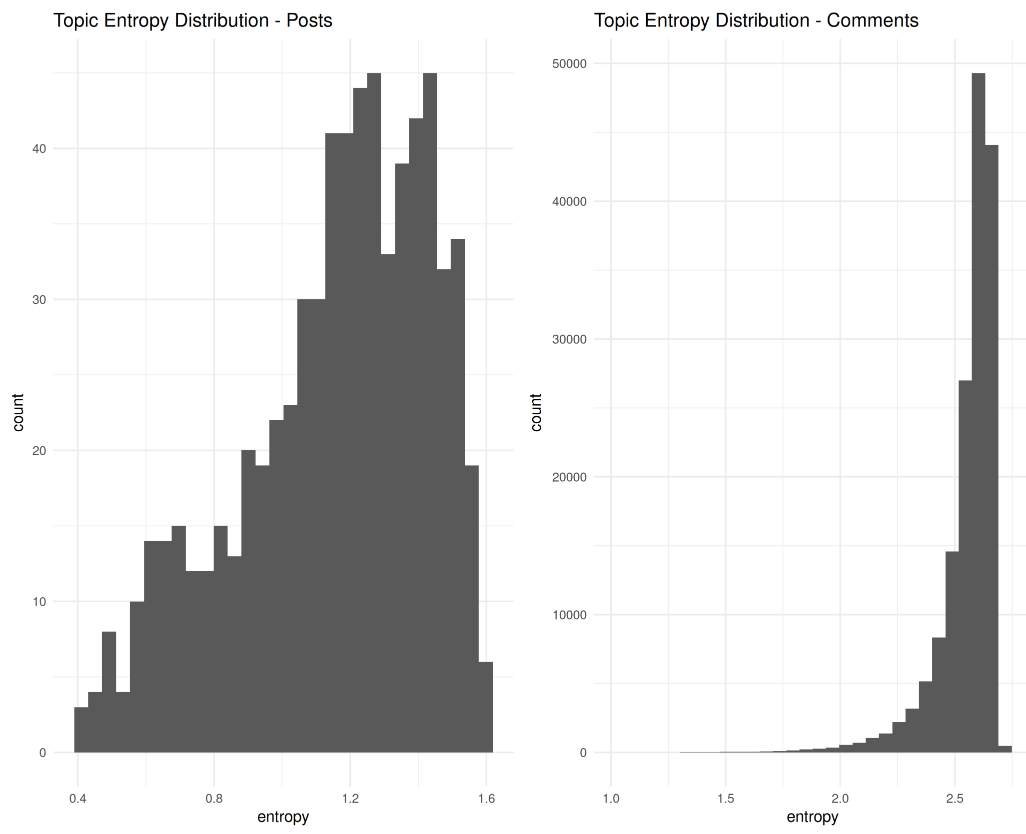

Figure 8: Topic Entropy Distributions for Media Posts and User Comments

The media posts exhibit a relatively normal distribution of entropy centred around $H \approx 1.2$, with a range of $[0.4, 1.6]$. This moderate entropy suggests that news articles typically blend multiple topics, reflecting journalistic practices of providing context and multiple angles on immigration issues. The symmetrical shape indicates a consistent level of topic mixing across the media corpus. In contrast, user comments display a markedly different distribution, heavily skewed towards higher entropy values ($H \in [1.0, 2.5]$) with a pronounced peak around $H \approx 2.4$. This substantially higher entropy indicates that comments tend to incorporate a broader mix of topics than media posts. The right-skewed distribution suggests that commenters frequently engage in discussions that bridge multiple aspects of the immigration debate simultaneously. The stark difference in entropy ranges ($\Delta H_{\max} \approx 0.9$) and distribution shapes between posts and comments quantitatively demonstrates their distinct discursive characteristics:

- Media posts maintain more focused thematic structures ($\sigma_H^2 \approx 0.2$), consistent with professional journalistic practices
- User comments exhibit greater topic integration ($\sigma_H^2 \approx 0.3$), reflecting the more conversational and interconnected nature of public discourse
- The minimal overlap between distributions indicates fundamentally different approaches to topic mixing in these two discourse spaces

These results are consistent with an observational IPR perspective in which comment threads display greater response-side dispersion than posts (Fig. 8). Comments exhibit a systematically higher topic–mixture entropy than posts (Fig. 8). While media coverage tends to maintain clearer topical boundaries, public discourse in comments more frequently synthesizes multiple aspects of the immigration debate within single contributions, indicating that audience discussion spans a broader set of themes within threads.

5.2.4. Auxiliary shared-space lexical divergence (echo vs. expansion)

Because posts and comments are modeled via separate STMs with potentially different topic dimensionalities, we complement topic-based comparisons with a shared-space, thread-level *lexical* divergence measure (Appendix A.7). Specifically, we compute the Jensen–Shannon distance between each post and the aggregate of its associated comments in the post-vocabulary space, which provides a well-defined symmetric notion of echo versus expansion at the thread level.

Token coverage of comments in the post vocabulary is 47.28% overall (range 43.68%–47.94% across media outlets). This quantity refers only to the shared lexical support induced by the post vocabulary in this auxiliary diagnostic and should not be interpreted as a loss of comment data from the main analysis. The resulting divergence is systematically high (median 0.786, IQR [0.765, 0.807]; 687/689 posts computable), indicating that comment threads extend well beyond the lexical footprint of the initiating post on the shared support. Given the marked asymmetry between the much smaller post corpus and the much larger comment corpus, this level of overlap is expected and is consistent with substantial response-side lexical expansion.

Outlet-level differences are present but moderate (median range 0.765–0.809), and the association with thread size is weak (Spearman $\rho \approx -0.096$). Accordingly, this measure should be read as a conservative diagnostic of overlap/divergence on restricted lexical support, rather than as a complete estimate of semantic mismatch between posts and comments. By construction, it cannot on its own distinguish whether the observed divergence reflects expansion, contestation, or reinterpretation on the comment side; it indicates only that comment-thread language goes beyond a simple lexical echo of the initiating post. Full distributions and outlet-stratified results are reported in Appendix A.7.

5.2.5. Media Effect on Media Post Topics Prevalence

To examine how different media outlets emphasize the different media post topics, we move to analyze the topic prevalence variations across media outlets. Figure 9 presents a heatmap visualization of topic prevalence by the media outlet, where effect sizes indicate the deviation from the baseline prevalence of each topic.

The analysis reveals distinct patterns in how different media outlets cover immigration-related topics. "European Politics" shows significant variation across media outlets, with *La Repubblica* ($\gamma = 0.297, p < 0.001$) and *Resto del Carlino* ($\gamma = 0.312, p < 0.001$) demonstrating the strongest positive associations. This suggests these outlets tend to frame immigration issues within a broader European political context.

"Sea Rescue Operations" exhibits strong positive associations with certain media outlets (*La Repubblica*: $\gamma = 0.451, p < 0.001$; *Resto del Carlino*: $\gamma = 0.285, p < 0.001$), while others demonstrate significant negative associations (*Il Fatto Quotidiano*: $\gamma = -0.148, p < 0.01$;*Fanpage*: $\gamma = -0.140, p < 0.01$).

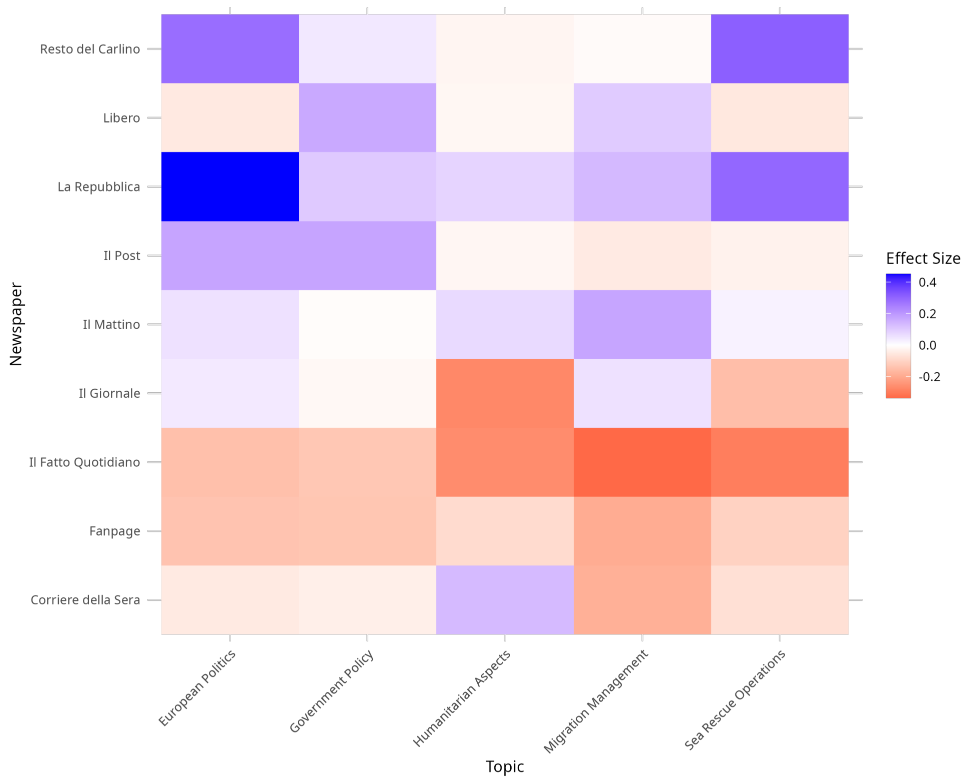

Figure 9: Media Outlet Effect On Posts Topic Prevalence. The colour indicates the effect sizes' deviation from the baseline prevalence of each topic; blue indicates a positive association (prevalence higher than the baseline), and red indicates a negative association (prevalence lower than the baseline).

The "Humanitarian Aspects" topic shows particularly negative associations with *Il Fatto Quotidiano* ($\gamma = -0.264, p < 0.001$) and *Il Giornale* ($\gamma = -0.274, p < 0.001$) while maintaining slightly positive associations with outlets like *Corriere della Sera* ($\gamma = 0.134, p < 0.05$) and *La Repubblica* ($\gamma = 0.082, p < 0.05$).

"Migration Management" demonstrates an interesting pattern of coverage, with *Il Mattino* showing the strongest positive association ($\gamma = 0.175, p < 0.01$), while *Il Fatto Quotidiano* shows the most negative association ($\gamma = -0.336, p < 0.001$). This suggests significant variation in how different outlets approach the operational aspects of migration policy implementation.

The "Government Policy" topic shows moderate variation, with *Il Post* demonstrating the most extensive coverage ($\gamma = 0.176, p < 0.01$) and Libero following closely ($\gamma = 0.168, p < 0.01$). This pattern indicates these outlets tend to focus more on policy-related aspects of the immigration discourse.

These findings are consistent with RQ1. They highlight the significant variations in how Italian media outlets frame and emphasize different aspects of immigration coverage, reflecting their editorial priorities and potentially their ideological orientations.

5.2.6. Media Effect on User Comment Topics Prevalence

We now discuss the effect of media outlets on the prevalence of comments' topics, as shown in Figure 10. This analysis reveals several distinctive patterns across different media outlets' commenting communities. The topic distribution shows variations that sometimes diverge from the media outlets' editorial stances.

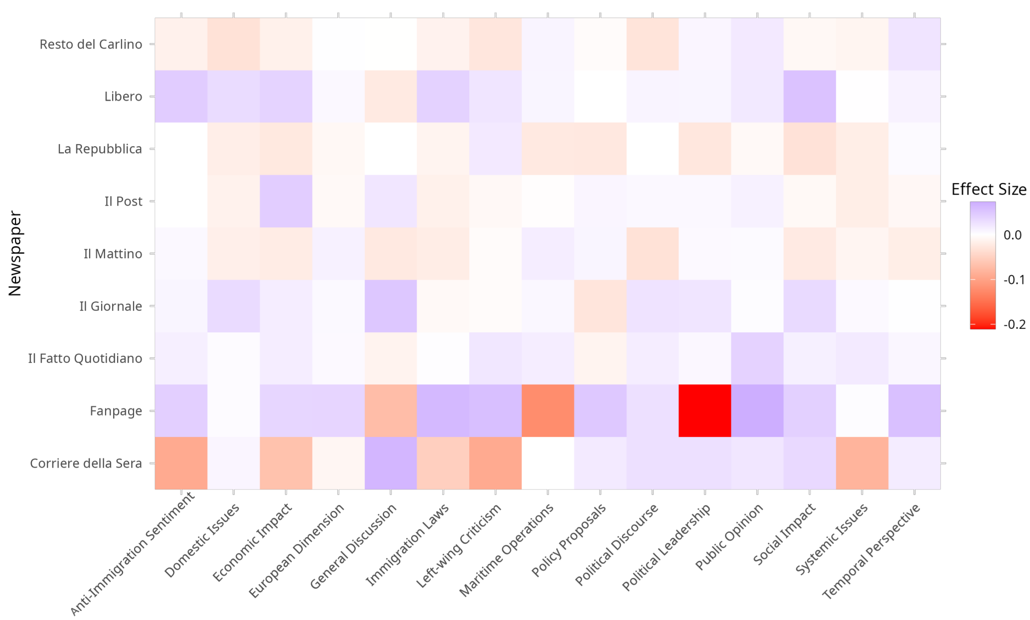

Figure 10: Media Outlet Effect on Comments Topic Prevalence. Colour indicates the effect sizes' deviation from the baseline prevalence of each topic; blue indicates a positive association (prevalence higher than the baseline), and red indicates a negative association (prevalence lower than the baseline).

Fanpage's comment section shows the most pronounced effects across multiple topics. It demonstrates strong negative associations with *Political Leadership* ($\gamma = -0.211$) and *Maritime Operations* ($\gamma = -0.123$) while showing positive associations with *Public Opinion* ($\gamma = 0.073$) and *Immigration Laws* ($\gamma = 0.064$). This suggests a commenting community more focused on public discourse and legal frameworks than operational or political aspects. *Corriere della Sera* exhibits interesting contrasts, with positive associations in *General Discussion* ($\gamma = 0.066$) and *Social Impact* ($\gamma = 0.034$) but negative associations with *Anti-Immigration Sentiment* ($\gamma = -0.092$) and *Left-wing Criticism* ($\gamma = -0.093$). This pattern suggests a more moderate, discussion-oriented commenting community. *Libero's* comment section shows consistent positive associations across multiple topics, particularly in *Social Impact* ($\gamma = 0.055$), *Anti-Immigration Sentiment* ($\gamma = 0.046$), and *Immigration Laws* ($\gamma = 0.041$). This pattern indicates a commenting community actively engaged in various aspects of the immigration debate, with a particular focus on social and legal implications. *Il Giornale* shows moderate positive effects in *Domestic Issues* ($\gamma = 0.033$) and *Social Impact* ($\gamma = 0.033$), suggesting a focus on local and social implications of immigration. *La Repubblica*'s comments show relatively weak effects across most topics, with slight negative associations in *Social Impact* ($\gamma = -0.032$) and *Political Leadership* ($\gamma = -0.027$). *Il Post* demonstrates the strongest positive association with *Economic Impact* ($\gamma = 0.045$), indicating a commenting community particularly focused on financial aspects. *Il Fatto Quotidiano* shows moderate positive effects in Public Opinion ($\gamma = 0.040$) and *Left-wing Criticism* ($\gamma = 0.022$).

These findings further strengthen the affirmative answer to RQ2 and suggest that different media outlets attract different commenting communities with their characteristic approaches to discussing immigration issues.

5.3. Regularized Post–Comment Associations (RQ3)

As explained in Section 4.3, association inference through a regularized regression approach allows us to quantify how media content relates to public discourse regarding migration. The regularization and bootstrapping procedures ensure that the results are robust to multicollinearity and sample variability, which is critical in high-dimensional settings such as topic-based inference. The effect matrix $\hat{\mathbf{B}}$ is visualized as a heatmap, where each cell's colour intensity represents the magnitude and direction of the effect of a specific post topic on a specific comment topic. In particular, colours range from dark blue, indicating a strongly positive impact, to pale pink, which denotes a strongly negative one. Significant effects are marked by an asterisk

(*), indicating confidence intervals that do not contain zero, thus providing a measure of the robustness of the estimated relationships.

5.3.1. Media Topics Influence on Comment Topics

First, we focus on the influence of media-post topics on user-comment topics. Figure 11 reveals the distinctive influence patterns across different media topics on user discussions.

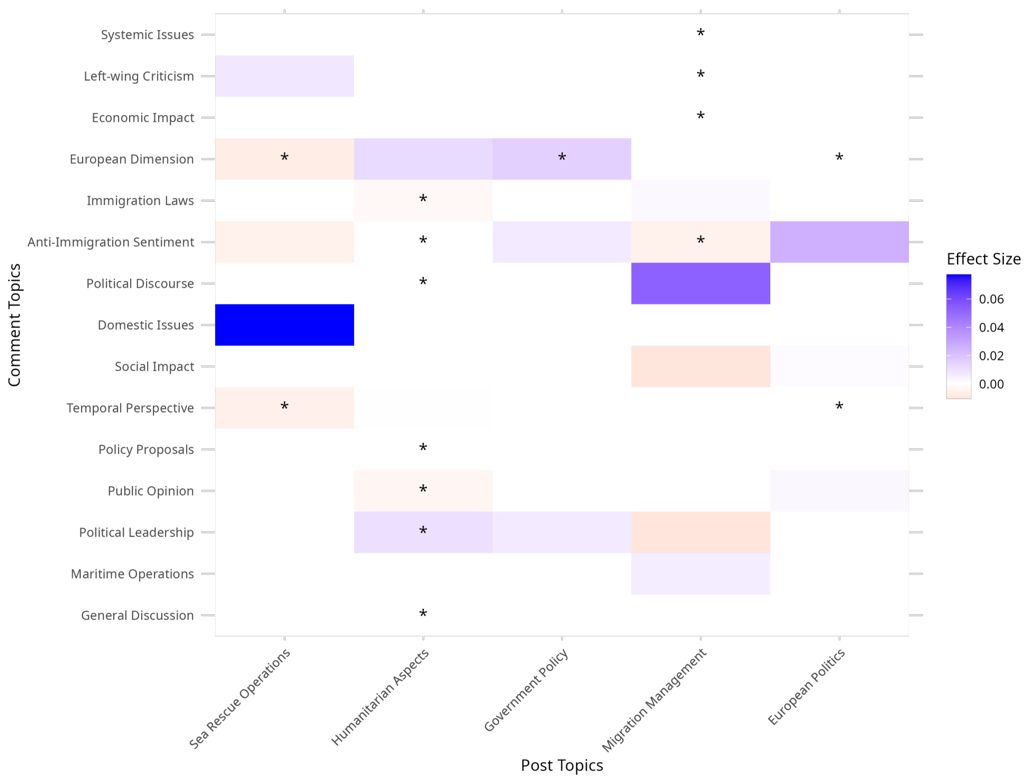

Figure 11: Regularized Effects of Post Topics on Comment Topics. Colors indicate effect size, with blue representing positive effects and pink representing negative effects. Asterisks (*) indicate significant effects based on 95% bootstrap confidence intervals.

Our analysis reveals relationships between specific media topics and public discourse patterns. Media posts focusing on *Humanitarian Aspects* demonstrate the widest influence, significantly affecting seven comment topics: *Immigration Laws*, *Anti-Immigration Sentiment*, *Political Discourse*, *Policy Proposals*, *Public Opinion*, *Political Leadership*, and *General Discussion*. The predominantly negative effects (indicated by pink colouration) suggest that humanitarian framing in media coverage may actually temper discussions in these areas, with the exception of *Political Leadership*, where a slightly positive effect is observed. This indicates that humanitarian narratives might encourage more engagement with leadership aspects of migration governance.

The *Migration Management* topic in media posts shows significant influence (indicated by purple to light blue colouration) on four comment topics: *Systemic Issues*, *Left-wing Criticism*, *Economic Impact*, with mostly slightly positive effect, and *Anti-Immigration Sentiment*, and *Anti-Immigration Sentiment* with a negative influence. This suggests that administrative and operational discourse in media tends to stimulate discussions about structural problems, critical perspectives from the left, and economic considerations, while anti-immigration viewpoints among users are discouraged.

Sea Rescue Operations in media content significantly affects *European Dimension* and *Temporal Perspective* comment topics, both with negative effects (light pink colouration). This indicates that coverage of maritime rescue activities may actually reduce discussion of broader European contexts and historical/future perspectives on migration, potentially narrowing the discourse to immediate operational concerns.

Similarly, *European Politics* as a media topic shows significant positive effects (purple to light blue colouration) on the same two comment topics: *European Dimension* and *Temporal Perspective*. This creates an interesting contrast with *Sea Rescue Operations*, as EU-level political discourse appears to expand rather than constrain the spatial and temporal frameworks for discussing immigration issues.

Finally, *Government Policy* media content shows a significant positive effect (light blue colouration) solely on the *European Dimension* comment topic. This suggests that discussions of national migration policies in media tend to broaden user discourse toward considering European-wide implications and contexts. Overall, research question RQ3 can be answered by the statistically significant topic-specific transmission from media posts to user comments. The findings reveal a complex interplay between media framing and public discourse, with varying directions and magnitudes of effects across different topical combinations.

This analysis underscores the nuanced role of media in structuring public discourse on migration: humanitarian framing tends to broaden discussion with tempering effects, whereas operational and policy-focused coverage stimulates targeted domains of engagement. Consistent with this, elastic-net estimates linking inputs to responses are modest and heterogeneous—outlet identity and post-topic composition explain limited but non-zero variation in comment-topic shares, with small effect sizes that vary across issues. Taken together, these patterns are consistent with differentiated audience-side interpretation and dispersion dynamics discussed in prior literature, including two-step flow, agenda melding, and event-driven surges in attention, without directly identifying the underlying mechanisms. Media organizations and policymakers should therefore communicate with these channeling dynamics in mind, recognizing that frames can systematically steer discourse rather than simply amplify it.

These conclusions are consistent with evidence that news emphasis and policy salience exert second-level agenda-setting effects on social-media discourse (Kim et al., 2025) and that right-leaning outlets can seed cross-media threat narratives about immigrants that travel across platforms (Zhang and Lian, 2025), all within comment arenas where user agency and cross-cutting exposure complicate one-to-one input→output transmission (Yarchi et al., 2024).

5.3.2. Media Outlet Influence on Comment Topics

We then focus on the influence of media outlets on comment topics, as presented in Figure 12. Our analysis reveals a complex pattern of relationships, with substantial variation in both the breadth and direction of influence across different Italian media outlets.

La Repubblica demonstrates the most extensive influence on user discourse, significantly affecting eleven different comment topics. It positively influences *Political Leadership*, *Anti-Immigration Sentiment*, *Temporal Perspective* and *European Dimension*, indicated by light blue coloration. In contrast, it shows significant negative effects (light pink coloration) on *Domestic Issues*. This pattern suggests that *La Repubblica*'s coverage tends to stimulate general immigration discussions and legal-institutional perspectives while potentially constraining domestic discussions.

Corriere della Sera significantly affects eight comment topics, with positive influence (light blue) on *Temporal Perspective* and *Left-wing Criticism*. It exhibits negative effects (pink coloration) on *European Dimension* and *Maritime Operations*, suggesting that its coverage may reduce the European framing of migration issues in favor of more nationally oriented perspectives.

Resto del Carlino significantly affects eight comment topics. Its positive effect is *European Dimension* (light purple). It shows negative effects

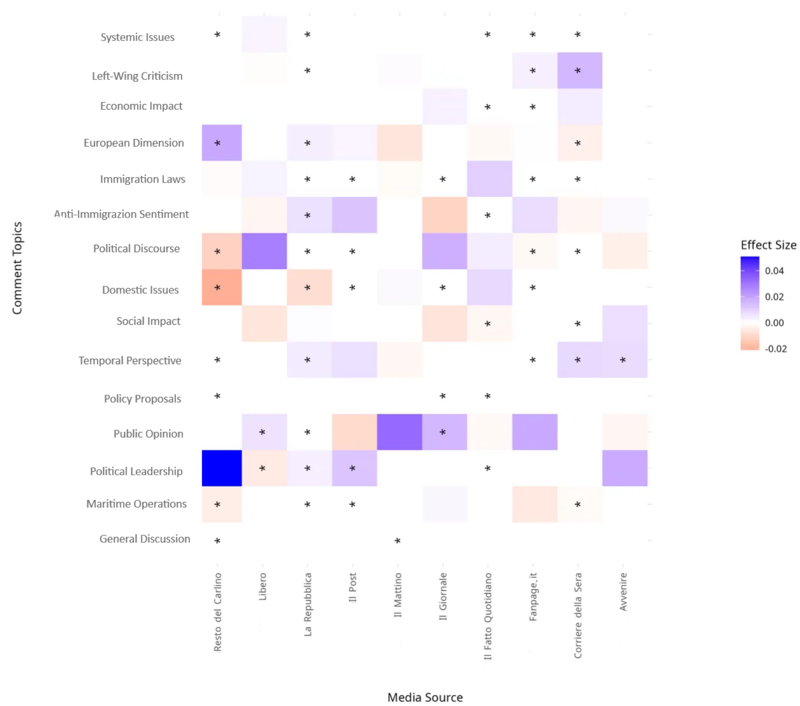

Figure 12: Media Outlet Effects on Comment Topics. Colors indicate effect size, with blue representing positive effects and red representing negative effects. Asterisks (*) indicate significant effects based on 95% bootstrap confidence intervals.

(pink colouration) on *Maritime Operations*, *Domestic Issues* and *Political Discourse*, suggesting its coverage may narrow discussion toward leadership aspects while constraining broader contextual discourse.

Fanpage.it similarly influences seven topics, with positive effects (light purple) on *Temporal Perspective*, *Anti-Immigration Sentiment*, *European Dimension*, *Left-wing Criticism*, and *Systemic Issues*. It negatively influences *Political Discourse* and *Immigration Laws*, potentially shifting the discussion from legal-political frameworks toward broader systemic and critical perspectives.

Il Fatto Quotidiano demonstrates significant positive effects (light purple) on *Left-wing Criticism* while slightly negatively influencing *Political Discourse*. This pattern suggests its coverage may amplify critical and sentiment-laden discussions while potentially constraining politically focused discourse.

Il Giornale significantly positively affects discussion on *Public Opinion* with a higher effect size, indicating a potential tendency to stimulate wide-range-driven discussion.

Il Post significantly and positively influences *Political Leadership* with higher, and *Maritime Operations*, *Domestic Issues* and *Immigration Laws* with a slightly positive effect. This pattern suggests Il Post may foster operational and country-level political discussions while tempering sentiment-driven discourse.

Libero shows significant positive effects on *Public Opinion* and a negative effect on *Political Leadership*, suggesting its potential to amplify public opinion while constraining leadership-focused discussions.

Il Mattino demonstrates a limited significant influence, with a positive effect on *General Discussion*, indicating a narrowly focused impact on general discourse.

Avvenire also shows a limited association, with only a positive effect on the *temporal perspective*, suggesting a minimal impact on the broader landscape of immigration discourse.

Cross-cutting patterns emerge when examining specific comment topics. *Anti-Immigration Sentiment* is positively influenced by *La Repubblica* and *Il Fatto Quotidiano*, suggesting this discourse may be amplified across these media ecosystems. Similarly, *European Dimension* receives positive influence from multiple outlets, *Resto del Carlino*, and a negative effect with *Il Corriere della Sera*, highlighting competing frames for continental versus national perspectives on migration.

While the effect magnitudes are relatively modest ($|\beta| < 0.05$), indicating that media outlet alone explains only a portion of the topic variation in comments, the patterns across outlets and topics reveal that specific characteristics of media outlets influence the contours of public discussion on immigration. These findings address RQ3 by demonstrating how different media outlets in Italy cultivate distinct patterns of public discourse on immigration, with implications for understanding how media pluralism affects the diversity and framing of societal debates on contentious issues.

5.4. Emotional Tone and Hostility Indices in Comments (RQ4)

We report a comparative sentiment analysis of user comments responding to Facebook posts across outlets. Indices and uncertainty are defined in Appendix A.5. Figure 13 shows a lollipop chart that jointly displays polarity (left) and intensity (right) by outlet, including 95% confidence intervals.

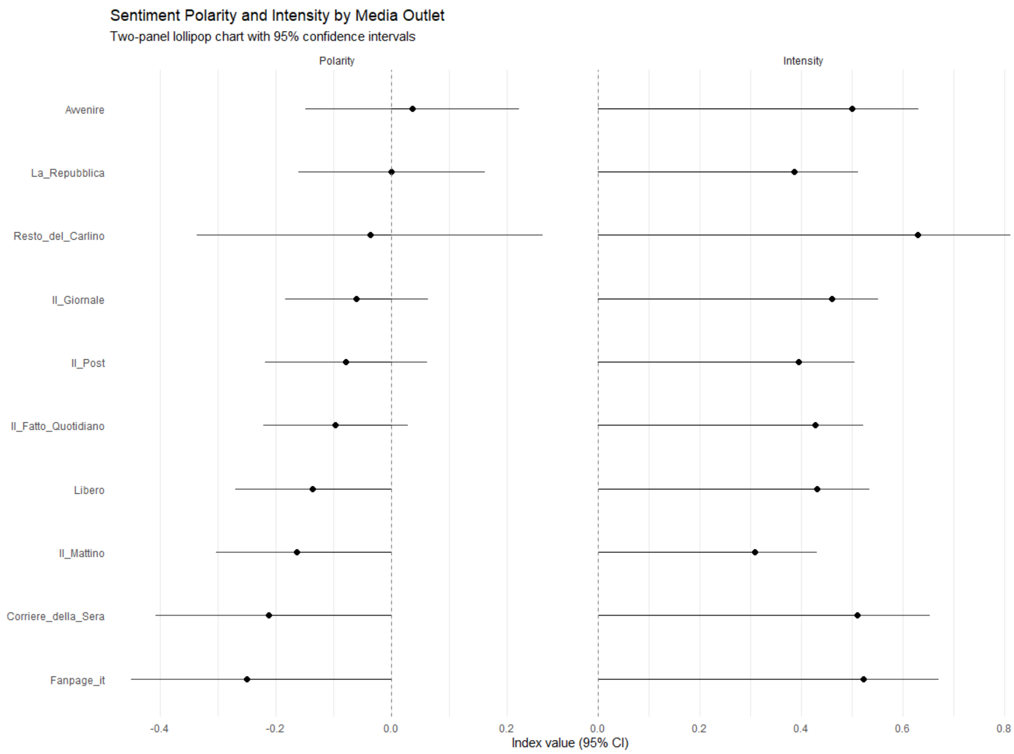

Figure 13: **Polarity and intensity of user comments by outlet (with 95% CIs).** Points are outlet-level means; horizontal lines are 95% confidence intervals (methods in Appendix A.5); stems are anchored at zero. Outlets are ordered by polarity. The left panel shows the Polarity Index (negative values indicate more negative-than-positive comments); the right panel shows the Intensity Index (share of non-neutral comments).

The left panel indicates that outlet-level polarity is generally non-positive, with heterogeneity across outlets. *Avvenire* lies closest to neutrality (slightly positive), whereas *Fanpage.it* and *Corriere della Sera* are among the most negative. Several confidence intervals overlap zero, indicating that departures from neutrality are modest for some outlets. The right panel shows moderate-to-high activation across outlets (Intensity Index typically around 0.4–0.55). *Resto del Carlino*, *Corriere della Sera*, and *Avvenire* sit at the upper end of the intensity distribution, while other outlets cluster closely, with overlapping intervals. Taken together, comment tone is predominantly negative but with varying strength of activation across outlets; between-outlet differences are present but often accompanied by overlapping uncertainty bands.

The analysis proceeds by providing a deeper focus on the nature of nega-

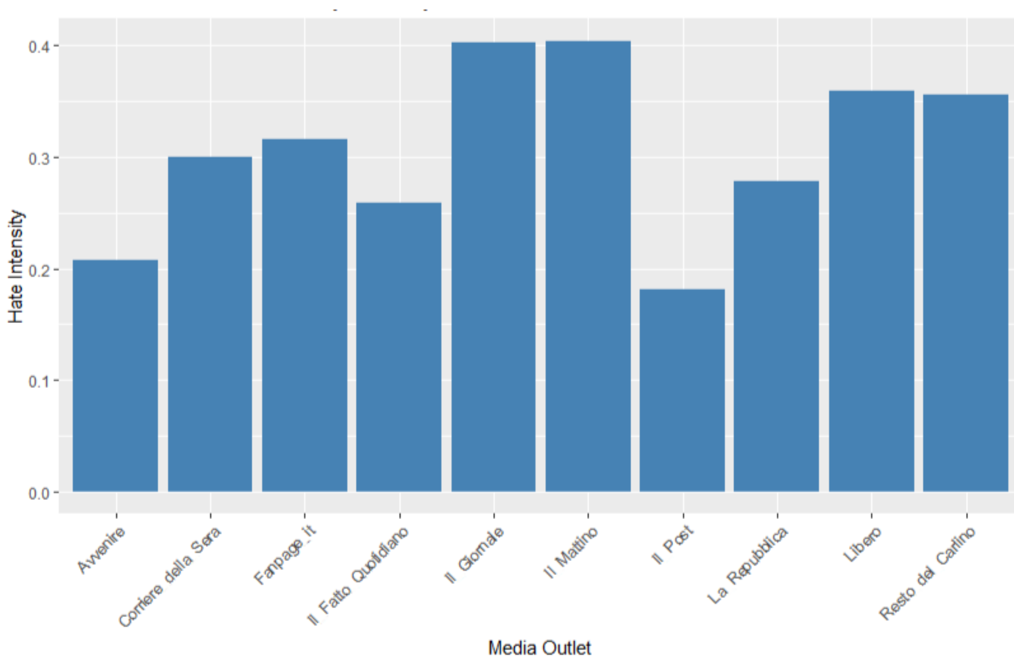

Figure 14: Bar chart showing hate intensity values across ten Italian media outlets. The y-axis represents the hate intensity index (ranging from 0 to approximately 0.4), while the x-axis displays the different media outlets. Each bar represents the average hate intensity measured in user comments for the corresponding outlet.

tive sentiment, specifically on hate intensity, quantifying how frequently and strongly hateful expressions appear in the comment sections of each outlet. As shown in Figure 14, *Libero*, *il Giornale*, *Il Mattino* and *Resto del Carlino* display higher mean hate intensity, pointing to a greater tendency toward hostile or hateful language in their comment sections. By contrast, *Corriere della Sera* and *Avvenire* show lower average hate intensity, suggesting that even though their overall sentiment may be negative, the proportion of explicitly hateful remarks is somewhat reduced. This distinction between general negativity and specific hate intensity provides important nuance to our understanding of the quality of discourse across different media environments.

Moving from media outlet comparisons to media topic comparisons, we examined hate polarization across the five media topics identified in the topic modeling analysis; results are shown in Figure 15. This perspective reveals how specific content themes, rather than just outlet identity, may influence

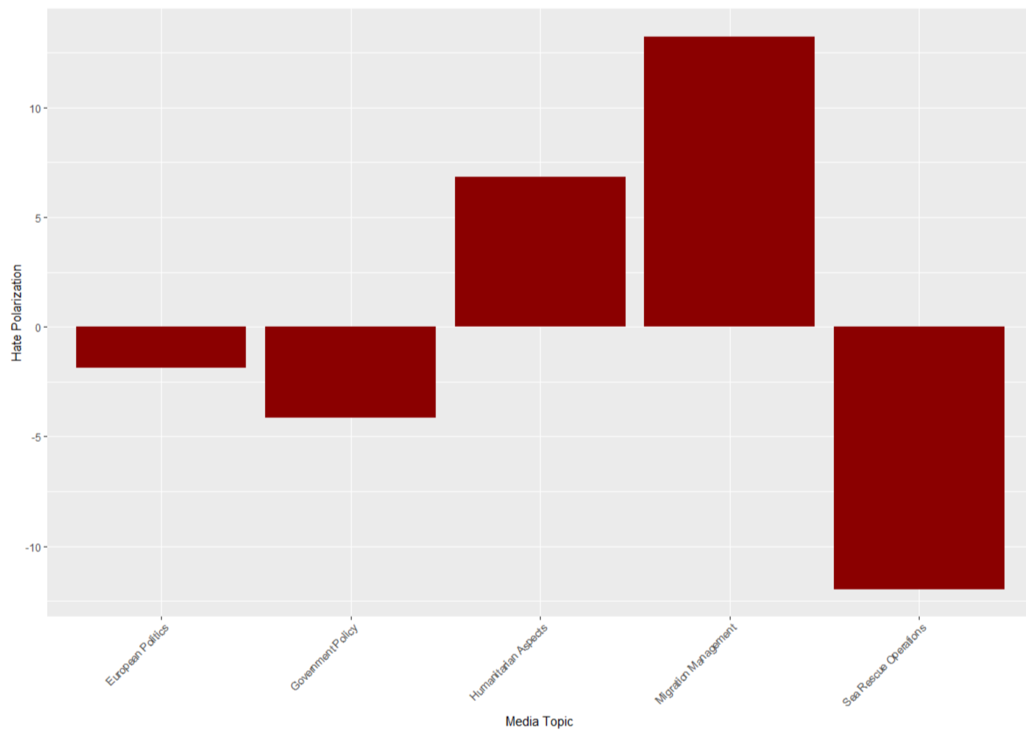

Figure 15: Bar chart showing hate polarization values across five media topics identified through topic modeling. Positive values indicate topics associated with higher levels of hate speech in user comments, while negative values represent topics with lower hate speech prevalence compared to the average.

the intensity of hateful discourse. *Migration Management* emerges with the highest polarization score, suggesting that discussions of this topic, which focuses on the operational aspects of handling migration flows, tend to elicit more intense hateful or hostile sentiments. In sharp contrast, *Sea Rescue Operations* shows a pronounced negative polarization, indicating that discussions around rescue activities generate comparatively less hateful discourse. *Europe Politics* and *Government Policy* are associated with relatively lower polarization scores. These findings highlight how the framing and subject matter of immigration coverage may independently influence the hostility of public responses, with operational and management aspects appearing to trigger more polarized reactions than humanitarian or policy-focused discussions. Therefore, concerning the research question RQ4, we find that different media outlets and topics influence variations in the emotional tone of public discussions.

Then, we move to the investigation of extreme hate speech. Figure 16 shows the results. *Il Giornale*, *Libero* and *Il Mattino* exhibit relatively higher shares of extreme hate comments, aligning with their elevated hate intensity scores identified in the previous figure. Outlets with lower hate intensity, such as *Il Post*, correspondingly show smaller proportions of extremely hateful content. This consistent pattern suggests a strong relationship between overall hate intensity and the fraction of comments that reach extreme levels of hostility, indicating that certain media environments may systematically foster more extreme expressions than others.

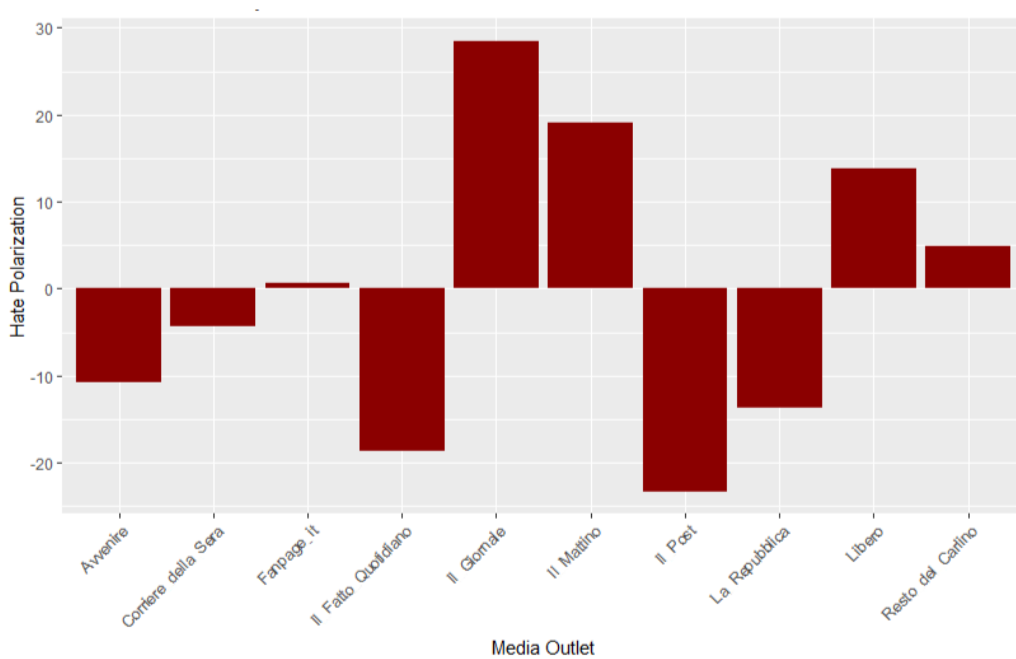

Figure 16: Bar chart depicting hate polarization values across ten Italian media outlets. Positive values indicate topics associated with higher levels of hate speech in user comments, while negative values represent topics with lower hate speech prevalence compared to the average.

When interpreted alongside topic modeling and association analysis findings, these sentiment analysis results provide a more complete picture of the media-public discourse dynamic. Media outlets that emphasize certain topics (such as Migration Management) shape what the public discusses and are associated with the emotional tenor of those discussions. The connection between topic emphasis, outlet identity, and hate speech intensity suggests that editorial choices about how to frame immigration issues may relate to downstream differences in the civility of public discourse. Outlets with more negative polarity but lower hate intensity demonstrate that negative sentiment toward immigration issues can be expressed without resorting to extreme hostility, pointing to potential strategies for fostering more constructive public engagement on contentious topics.

Overall, these findings reveal that while user comments on immigration-related posts are generally negatively oriented across all outlets, the degree of negativity and prevalence of hateful content varies significantly. Media environments with higher emotional intensity are more likely to feature overtly hostile language, whereas outlets with lower emotional intensity tend to host negative sentiment expressed in less extreme forms. This pattern aligns with our earlier observations about topic prevalence and suggests that both what is discussed (topic) and where it is discussed (outlet) shape the emotional quality of public engagement with immigration issues. These insights not only enhance our understanding of online discourse dynamics but also point to potential intervention points for media organizations seeking to foster more constructive public discussions on sensitive social issues.

6. Conclusions

6.1. Discussion of the main findings

This study analyzed associations between media narratives and public discourse surrounding immigration in Italy, focusing on how topic structure, lexical divergence, and affective tone vary between media outlet posts and user-generated comments on social platforms. Within the *input–processing–response* (IPR) framework, outlet posts constitute observable *inputs*, comment threads are observable *responses*, and audience cognition and interaction are treated as an interpretive *processing* layer rather than as directly measured constructs. The main contributions and findings can be summarized as follows.

Contributions. Beyond providing a detailed description of media and audience patterns, this study contributes to computational research on migration-related communication and online hostility in three main respects. First, we jointly analyse Facebook posts and their corresponding user comments within the same discussion threads across multiple Italian outlets, which allows us to quantify the extent to which topical emphases in media inputs are echoed, transformed, or re-weighted in audience responses. This extends prior work on immigration sentiment and communication on Twitter and Facebook, which typically models posts or comments in isolation or at the level of outlet or leader aggregates (Bartlett and Norrie, 2015; Rowe et al., 2022; Menshikova and van Tubergen, 2022; Prislei et al., 2022). Second, we operationalize the IPR perspective as a thread-level observational design that links posts as inputs to associated comment threads as observable responses, and use regularized association analyses to estimate how specific media topics and outlet identities relate to comment-topic distributions. This goes beyond topic-model applications that rarely model post–comment mappings explicitly in migration and hate-speech research (Viola and Verheul, 2020; Calderón et al., 2020; Pierri, 2024; Zhang and Lian, 2025). Third, by integrating topic structure with transformer-based sentiment and hate-speech indices, we connect media-framing and media-effects traditions (McCombs and Shaw, 1972; Entman, 1993; Helbling, 2014) with computational studies of online sentiment and hostile speech (Fortuna and Nunes, 2018; Burnap and Williams, 2015, 2016; Garland et al., 2022). Within this design, processing is treated as an interpretive layer rather than as a directly measured mechanism, making it possible to relate specific frames and topics not only to what audiences discuss, but also to the emotional tone and hostility associated with those discussions.

Findings. Salience and tone have been analysed using the *Fightin' Words* methodology. We found that outlet posts about immigration vary systematically across different media outlets, with each group favouring particular lexical choices. The posts reveal substantial lexical variation tied both to editorial orientation (humanitarian vs. alarmist frames), local vs. national focus (municipal concerns vs. large-scale narratives), and the degree of policy-focused discourse. These distinctions indicate that the same topic (immigration) is packaged with qualitatively different language patterns depending on the outlet environment. Structural Topic Modeling further indicates substantial variation in thematic structure across outlets. These findings highlight how Italian media outlets frame and emphasize different aspects of immigration coverage, reflecting editorial priorities and likely ideological orientations (see RQ1).

Moreover, our findings suggest that user comments do not merely replicate outlet narratives. Instead, they introduce additional topics and perspectives, and often surface polarized viewpoints. Consistent with prior literature, we identify distinct thematic structures between outlet posts and user comments, indicating that while coverage maintains relatively focused thematic boundaries, public discourse exhibits greater diversity and fragmentation (see RQ2). Within the IPR design, regularized association analyses further indicate that some post topics are systematically associated with subsequent comment-topic proportions and that these associations vary across outlets (see RQ3). Because the design aligns the media post as the unit of exposure with the associated comment thread as the observable response, these results can be interpreted as structured input–response associations rather than as simple co-occurrences between separate corpora. This approach allows us to identify post topics and outlet identities that are more strongly associated with specific discussion themes.

Finally, the analysis indicates that both outlet identity and specific topics are associated with variations in the emotional tone of public discourse (see RQ4). We observe that average sentiment polarity and intensity differ across outlets. Polarity values tend to be predominantly negative, while intensity levels suggest a moderate to high degree of emotional activation in user comments. Regarding hate intensity, the overall findings indicate that the index varies by outlet. The hate intensity indices suggest that some outlets are associated with higher levels of hostile or hateful language. These hate-related indicators should, however, be interpreted as coarse aggregate-level exploratory descriptors, calibrated through the in-domain validation reported in Appendix A.5.1, and not as standalone estimates of comment-level or absolute hate-speech prevalence.

These patterns, considered alongside the observed polarization of hate in user discourse, point to a co-variation between overall hate intensity and the share of comments reaching extreme levels of hostility. The hate polarization index by media topic further shows that certain immigration-related themes are more frequently associated with polarized and hostile responses than others. In particular, the presence of extreme hate speech appears more pronounced in discussions related to Humanitarian Aspects and Migration Management, indicating that some framings of immigration are linked to elevated hostility in public discourse. Overall, the hate speech analysis suggests that while user comments often align with the negativity present in media narratives, the intensity and polarization of hateful discourse vary across topics and media outlets.

Altogether, our findings provide a comprehensive description of how Italian media narratives and public discourse on immigration are linked within a post–thread observational design informed by the IPR perspective. The results are consistent with media outlets acting as agenda-setting inputs and participating in amplification dynamics, while comment-thread responses show divergence in topic emphasis and emotional tone that is compatible with differentiated audience-side interpretation. User-generated discourse does not passively mirror outlet content; rather, it exhibits structured input–response associations, particularly in emotional tone and in the salience of polarizing narratives. The findings on hate speech are especially salient: certain outlets and topics are disproportionately associated with extreme hostility, especially around humanitarian issues and migration governance, underscoring the complexity of the media–audience relationship in this domain.

6.2. Policy implications

In this section we provide and discuss some implications with the caveat that our analyses are observational and associational rather than causally identified. The findings of this study could nonetheless inform the management of migration-related communication across media regulation, platform governance, and public communication strategies. In particular, the hate intensity and polarization indices show that discussions centred on *Humanitarian Aspects* and *Migration Management* are systematically more likely to host extreme hostility than other topics, indicating that some frames of immigration are structurally exposed to heightened levels of antagonistic and dehumanizing speech.

First, these frame-specific patterns motivate attention to responsible editorial practices. Regulatory bodies and journalistic associations could promote guidelines that explicitly recognize that even humanitarian or policy-management coverage of immigration may trigger disproportionate hostile reactions in comment sections. This supports media outlets routines that combine accurate, evidence-based reporting with proactive risk assessment of headlines, visuals, and narrative emphasis, especially for stories that foreground rescue operations, victimization, or conflicts around migration governance.

Second, social media platforms might invest in content-moderation strategies that are sensitive not only to explicit hate speech, but also to the topics and frames that, in empirical analyses, tend to be associated with spikes in hostility. Frame-aware tools that flag high-risk discussions—such as threads on humanitarian crises or migration-management disputes—and monitor user reactions in real time could help moderators prioritize resources, introduce friction (for example, prompts before posting), or deploy counter-speech in contexts where escalation is more likely.

Third, public institutions and civil society organizations may prioritize media literacy initiatives that help citizens critically evaluate emotionally charged narratives around humanitarian emergencies and migration policy. Strengthening the capacity to recognize sensationalist or polarizing framings can reduce the impact of ideologically driven messaging and limit the diffusion of hostile discourse. Finally, policy interventions that support pluralism in immigration-related coverage—by funding independent journalism and fostering dialogue among diverse perspectives—could contribute to a public sphere that remains inclusive even when addressing highly sensitive topics, thereby mitigating the risks of polarization associated with partisan media narratives.

6.3. Limitations and future research

Despite its contributions, this study has several limitations that also suggest directions for future work. One limitation is that the analysis focuses solely on Italy and on a single reference year. Both aspects offer opportunities for further investigation. The lack of territorial differentiation prevents assessment of how contextual factors shape the relationship between outlet posts and user comments. In Italy, social, economic, and political differences at regional and provincial levels likely influence public attitudes toward migrants and may affect the tone and content of online comments. Likewise, restricting the analysis to a single year does not allow examination of temporal trends, which often carry relevant information for understanding complex phenomena and their implications.

Both limitations may relate to another feature of our results, namely the relatively small estimated coefficients ($|\beta| < 0.05$). These magnitudes indicate that outlet and topic variables, while informative, account for a limited portion of the variation in public-discourse patterns. This suggests a more complex relationship between outlet content and user comments than our models capture, potentially involving unmeasured variables such as audience characteristics, platform dynamics, or broader socio-political contexts and indicators. The small coefficients also warrant caution when assessing practical significance, as statistical significance does not necessarily imply substantial real-world impact. Future research could address these limitations by incorporating additional explanatory variables, modeling platform- and audience-level heterogeneity, and, where appropriate, employing designs that enable stronger causal interpretation (e.g., experiments, quasi-experiments, instrumental-variables strategies). Mixed-methods approaches—combining quantitative analyses with qualitative examinations of conversation dynamics—could also illuminate the mechanisms through which media inputs, audience processing, and discursive responses interact within the IPR framework.

Another limitation is related to unobserved confounding. We do not control for outlet audience size, page-level moderation, or exogenous shocks; these may influence both coverage and audience tone. The absence of temporal fixed effects further limits causal interpretation. These constraints, together with small effect sizes, motivate our observational and associational approach. More broadly, they reinforce the role of the IPR perspective in this study as an observational design for organizing input–response associations, rather than as a framework for identifying the causal mechanisms underlying audience processing.

List of Abbreviations

DTM Document–Term Matrix.

EU European Union.

GRINS Growing Resilient, INclusive and Sustainable (project).

HIIm Hate Intensity Index by media.

HPIm Hate Polarization Index by media.

HPImp Hate Polarization Index by media topic.

ID Identifier.

IIm Intensity Index by media.

IPR Input–Processing–Response (framework).

LLMs Large Language Models.

PIm Polarity Index by media.

RQ1–RQ4 Research Question 1 to Research Question 4.

S–O–R Stimulus–Organism–Response (framework).

STM Structural Topic Modeling.

URL(s) Uniform Resource Locator(s).

7. Declarations

Not applicable

8. Ethics approval and consent to participate

Not applicable

9. Consent for publication

Not applicable

10. Availability of data and material

Code used is publicly available in this repository. Posts and Comments cannot be made publicly available for reasons of privacy, but an anonymized version will be made available upon request.

11. Competing Interests

Authors have no competing interests as defined by Springer, or other interests that might be perceived to influence the results and/or discussion reported in this paper.

12. Funding

This study was funded by the European Union - NextGenerationEU, Mission 4, Component 2, in the framework of the GRINS -Growing Resilient, INclusive and Sustainable project (GRINS PE00000018 – CUP F53C22000760007). The views and opinions expressed are solely those of the authors and do not necessarily reflect those of the European Union, nor can the European Union be held responsible for them.

13. Authors' Contributions

M.O. and G.C. and C.M contributed to the conceptualization. M.O. and G.C. contributed to the methodology. C.M and S.T and F.M. supervised the manuscript. All authors wrote the main manuscript text. All authors reviewed the manuscript.

14. Acknowledgements

Appendix A. Mathematical and Computational Details

Appendix A.1. Preprocessing and Tokenization

We apply identical pipelines to posts and comments: lowercasing; URL normalization; punctuation stripping; stopword removal (Snowball + domain-specific list); lemmatization; frequent-collocation bigrams (min-count threshold); document trimming by minimum length and document-frequency bounds. Let $\mathcal{V}$ be the resulting vocabulary. All stochastic components (e.g., bigram mining, STM initialization) are seeded for exact reproducibility (see Section 10).

Appendix A.2. Structural Topic Modeling

Our analysis of immigration discourse in Italian media and social media comments relies on Structural Topic Modeling (STM), which extends beyond traditional topic modeling methods. We selected STM for its unique ability to incorporate contextual information—such as the source of news articles or their relationship to comments—directly into the topic estimation process. This feature makes it particularly valuable for our comparative analysis between media narratives and public responses. When applying STM to our corpus, we treat media posts and user comments as distinct document sets while maintaining their relational structure. Each document contains not only its textual content but also important metadata, such as which media outlet published a given article or which specific post a comment responds to. This approach allows us to trace thematic patterns across the media-public discourse pipeline. The technical foundation of STM involves a probabilistic framework where topics emerge as patterns of co-occurring words. Unlike simpler approaches, STM can model how document characteristics (like media outlets) influence topic prevalence. This means we can examine whether certain media outlets tend to emphasize particular aspects of immigration and, subsequently, whether these emphases carry over into user discussions. Before introducing the methodological details, we provide the required definitions of notation. We denote D to represent our corpus of documents, divided into posts (D_p) and comments (D_c). For each document $d \in D$, we observe words $w_{d,n}$ where $n \in 1, ..., N_d$, and document-level metadata X_d (media outlet for posts, and both media outlet and parent post ID for comments). The generative process for each document follows. Draw topic proportions from $\theta_d | X_d\gamma, \Sigma \sim \text{LogisticNormal}(X_d\gamma, \Sigma)$. For each word, draw a topic assignment from $z_{d,n}|\theta_d \sim \text{Multinomial}(\theta_d)$. Draw a word from $w_{d,n}|z_{d,n}, \beta_{1:K} \sim \text{Multinomial}(\beta_{z_{d,n}})$, where K is the number of topics, γ represents coefficient vectors for topic prevalence, and $\beta_{1:K}$ represents topic-specific word distributions.

Appendix A.2.1. Model selection and Preprocessing

Before model estimation, we implement a preprocessing pipeline to prepare the textual data for subsequent steps. First, we remove special characters, URLs, and standardized text encoding (Text cleaning); we then eliminate Italian stopwords and custom domain-specific stopwords (Stopword removal), and finally, we removed documents with fewer than 5 tokens (Document filtering). STM defines the number of topics as a model parameter; thus, the optimal number of topics K was selected through a grid search over $K \in 2, ..., 30$, evaluating Eq. A.1, where $\alpha = 0.5$ represents the trade-off between coherence and exclusivity.

$$\begin{aligned} K^* = \arg\max_K \Big[& \alpha \cdot \text{semantic coherence}(K) \\ & + (1-\alpha) \cdot \text{exclusivity}(K) \Big] \end{aligned} \tag{A.1}$$

Appendix A.2.2. Comparative Analysis Framework

To compare topic prevalence $\theta_{d,k}$ among the media outlets, we derived a comparative framework to estimate the effects of media outlets (measured by γ_k). Let $\theta_{d,k}$ represent the proportion of topic k in document d. To analyze media outlet effects on topic prevalence, we estimate:

$$\mathbb{E}[\theta_{d,k}|X_d] = \text{logistic}(\mu_k + X_d\gamma_k) \tag{A.2}$$

where $Xd = [\text{outlet dummies}, \text{post id dummies}]$ represents covariates for comment (both media outlet and post id), μ_k is the topic-specific intercept and γ_k represents the effects of document covariates on topic k. For each topic

k, we estimate media outlet effects through:

$$\hat{\gamma}j, k = \mathbb{E}[\theta d, k | X_d = j] - \mathbb{E}[\theta_{d,k} | X_d = \text{baseline}] \tag{A.3}$$

where j indexes media outlets and "baseline" represents a reference media outlet.

For hypothesis testing regarding media outlet effects, we employ the following test statistic:

$$T_{j,k} = \frac{\hat{\gamma}j, k}{\text{SE}(\hat{\gamma}j, k)} \sim t_{n-p} \tag{A.4}$$

where n is the number of documents and p is the number of parameters. We use the Benjamini-Hochberg procedure for multiple testing correction with a false discovery rate $\alpha = 0.05$.

Appendix A.2.3. LLM-based Topic Interpretation and Labeling

To enhance the interpretability of the Structural Topic Model results and ensure consistent topic labeling, we implemented an automated labeling approach using Large Language Models (LLMs). While traditional topic modeling produces collections of weighted keywords, interpreting these word clusters and assigning meaningful topic labels typically requires subjective human judgment, which can introduce inconsistency and researcher bias.

The automated topic labeling procedure consisted of several sequential steps. First, for each topic $k \in \{1, 2, ..., K\}$ identified in our Structural Topic Model, we extracted the top-N most probable words $w_{k,1}, w_{k,2}, ..., w_{k,N}$ along with their corresponding probability weights $\beta_{k,1}, \beta_{k,2}, ..., \beta_{k,N}$. This allowed us to capture the most representative words for each topic. Next, we constructed a standardized prompt for the LLM that contained context about the research domain (Italian media coverage of immigration), the specific corpus type (either media posts or user comments), the full list of weighted keywords for the topic, and instructions to generate a concise, descriptive label based solely on the provided keywords.

Upon receiving this prompt, the LLM generated a concise label and justification for each topic, which was then reviewed by the research team for validity. In cases where the automatically generated labels were ambiguous or required refinement, we implemented an iterative process. This process involved appending the initial label and justification to the prompt, providing additional context about similar topics, and instructing the LLM to revise the label with greater specificity.

We implemented several validation procedures to ensure the automated labeling process's reliability. First, we assessed multiple-generation consistency by generating three independent labels for each topic using slightly varied prompts and evaluating their semantic consistency. This approach helped identify topics with unstable interpretations that might require additional review. Second, we employed human verification, wherein humans independently reviewed a subset of automatically generated labels to assess the direct alignment between human experts and machine interpretations. This methodological approach allowed us to maintain consistent topic interpretation across both media posts and comment topics, reducing subjective bias in the labeling process while preserving the nuanced understanding of topic content. The resulting topic labels were then used throughout the analysis to facilitate clear communication of findings and enable meaningful comparison between media and public discourse patterns.

Appendix A.3. Association Analysis: Elastic-Net Regression

To investigate the association effect of media topics on public discourse, we employ a regularized regression approach based on an elastic net regression to estimate the influence of specific post topics (denoted by $\mathbf{X}$) on user comment topics (denoted by $\mathbf{Y}$). Let $\mathbf{X} \in \mathbb{R}^{n \times p}$ denote the predictor matrix, where each column represents a topic in media posts (post topics), and each row corresponds to an observation of topic proportions in a specific post. Similarly, let $\mathbf{Y} \in \mathbb{R}^{n \times q}$ denote the response matrix, where each column represents a topic in user comments (comment topics), and each row corresponds to an observation of topic proportions in comments related to the corresponding posts. In the following sections, we describe the estimation methodology and the significance estimation.

Appendix A.3.1. Regularized Regression for Effects Estimation

In order to estimate the effects of media outlets and media topics on comment topic j ($j = 1, \ldots, q$), we fit an elastic net regression model. The elastic net model combines the properties of the Lasso and Ridge regressions, balancing variable selection and coefficient shrinkage. For each response vector $\mathbf{y}_j \in \mathbb{R}^n$, we solve the optimization problem in Eq. A.5, where $\hat{\boldsymbol{\beta}}_j \in \mathbb{R}^p$ represents the estimated effect of each post topic on comment topic j, λ is the regularization parameter that controls the degree of shrinkage, and α is the mixing parameter controlling the balance between Lasso (ℓ_1) and Ridge (ℓ_2) penalties.

$$\hat{\boldsymbol{\beta}}_j = \arg\min_{\boldsymbol{\beta}_j} \Big\{ \frac{1}{2n} \big\| \mathbf{y}_j - \mathbf{X}\boldsymbol{\beta}_j \big\|_2^2 + \lambda\Big(\alpha \big\|\boldsymbol{\beta}_j\big\|_1 + \tfrac{1-\alpha}{2} \big\|\boldsymbol{\beta}_j\big\|_2^2 \Big) \Big\} \tag{A.5}$$

To select an optimal value for λ, we employ cross-validation to minimize the prediction error. The model is fit using the `glmnet` function, and the coefficients at $\lambda_{1\text{se}}$ (the value of λ within one standard error of the minimum cross-validation error) are used as our final estimates for each comment topic j. The estimated coefficients $\hat{\boldsymbol{\beta}}_j$ for each comment topic j are assembled into an effect matrix $\hat{\mathbf{B}} \in \mathbb{R}^{p\times q}$, where each element $\hat{B}_{ij}$ represents the estimated association effect of post topic i on comment topic j. Formally, we have:

$$\hat{\mathbf{B}} = \begin{bmatrix} \hat{\beta}_{11} & \hat{\beta}_{12} & \dots & \hat{\beta}_{1q} \\ \hat{\beta}_{21} & \hat{\beta}_{22} & \dots & \hat{\beta}_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\beta}_{p1} & \hat{\beta}_{p2} & \dots & \hat{\beta}_{pq} \end{bmatrix} \tag{A.6}$$

Appendix A.3.2. Bootstrap Confidence Intervals and Significance Testing

To assess the significance of the estimated effects, we employ a nonparametric bootstrap approach to construct confidence intervals for each element in $\hat{\mathbf{B}}$. We perform B bootstrap iterations. In each iteration $b = 1, \dots, B$, we resample the rows of $\mathbf{X}$ and $\mathbf{Y}$ with replacement to generate bootstrap samples $\mathbf{X}^{(b)}$ and $\mathbf{Y}^{(b)}$. For each bootstrap sample, we re-estimate the effect matrix $\hat{\mathbf{B}}^{(b)}$.

The 95% confidence interval for each effect $\hat{B}_{ij}$ is computed by taking the 2.5% and 97.5% percentiles of the bootstrap distribution in Eq. A.7.

$$\text{CI}_{ij} = \big[\text{quantile}\big(\{\hat{B}_{ij}^{(b)}\}_{b=1}^{B},\, 0.025\big),\ \text{quantile}\big(\{\hat{B}_{ij}^{(b)}\}_{b=1}^{B},\, 0.975\big) \big] \tag{A.7}$$

We consider an effect $\hat{B}_{ij}$ to be statistically significant if the confidence interval CI_{ij} does not contain zero.

Appendix A.4. Lexical Salience: Fightin' Words

The Fightin' Words method Monroe et al. (2008) is a statistical technique used to identify words that are significantly associated with different groups in a corpus. It is particularly effective in the context of sentiment analysis and textual comparison across different categories. The method is based on a Bayesian framework that models word frequencies in different corpora and estimates the deviation from expected distributions using a log-odds ratio with an informative prior.

Given a document-term matrix (DTM) representing a corpus, where documents are categorized into groups, Fightin' Words computes the log-odds ratio for each term across groups, adjusting for corpus-level word frequency. Formally, the method computes:

$$\delta_{g,w} = \log\left(\frac{p_{g,w}}{1-p_{g,w}}\right) - \log\left(\frac{p_{\neg g,w}}{1-p_{\neg g,w}}\right) \tag{A.8}$$

where $p_{g,w}$ is the probability of word w appearing in group g, and $p_{\neg g,w}$ is the probability of word w in all other groups. A Bayesian prior is incorporated to stabilize estimates when word counts are sparse, preventing overfitting due to rare words.

The standard deviation of $\delta_{g,w}$ is computed to derive a z-score ($\zeta_{g,w}$), which indicates the statistical significance of word usage differences between groups. Terms with high absolute ζ values are considered the most distinctive for each group.

We start by constructing a sentiment lexicon for Italian using the Sentix dataset. Words in the lexicon are aggregated and assigned sentiment scores based on their average positive and negative sentiment values. A word is classified as positive if its average positive score is greater than its negative score and vice versa.

A comprehensive stopword list combines standard Italian stopwords from the Snowball stemmer with additional custom stopwords. The text preprocessing pipeline involves: i) Lowercasing all text, ii) Removing punctuation and numeric values, iii) Eliminating extra spaces, iv)Removing predefined stopwords. This step ensures a clean and normalized corpus for subsequent analysis.

The corpus is transformed into a document-term matrix (DTM), representing the frequency of terms across documents. We then compute the log-odds ratios for each term across different media outlets. First, zero-frequency terms are removed, and weight normalization is applied to term frequencies. A prior distribution is estimated from the corpus, and posterior probabilities are computed by incorporating the prior. The most distinctive terms for each media outlet are extracted selecting the highest-ranked words based on their Fightin' Words z-scores.

The results are visualized using custom plotting that colours words based on their sentiment classification. Words are displayed based on their z-score magnitude, highlighting those most characteristic of each media outlet. This provides an intuitive way to interpret linguistic biases across media outlet and assess the sentiment-driven differences in media discourse.

Appendix A.5. Sentiment And Hate Speech Analysis

The fourth and final step of our investigation involves a comparative sentiment analysis of user comments posted on Facebook posts by media outlets on the topic of immigration. To this end, we define five indices aimed at evaluating the polarity, intensity, and hate speech characteristics of user-generated responses. Similar sentiment and hate speech quantification methods have been previously explored in online discourse analysis (Pang and Lee, 2008; Tetlock et al., 2008; Liu, 2012; Wilson et al., 2005; Waseem and Hovy, 2016; Davidson et al., 2017; Agresti, 2002; Conover et al., 2011). However, our approach introduces a novel set of indices tailored to analyze the role of media outlets in shaping sentiment and hate speech trends across topics.

More specifically, we introduce four indices at the media outlet level and one index at the media topic level:

1. **Polarity Index by media (PI_m)**: Measures the average sentiment polarity in user comments across different media outlets. While sentiment polarity indices have been widely used in literature Tetlock et al. (2008), our metric is uniquely designed to compare sentiment tendencies across media platforms, following a similar comparative approach as in Pang and Lee (2008).
2. **Intensity Index by media (II_m)**: Quantifies the degree of emotional activation in user comments. Unlike prior studies that model sentiment intensity as a continuous score (Wilson et al., 2005; Liu, 2012), we introduce a comparative index to assess emotional engagement per media outlet.
3. **Hate Intensity Index by media (HII_m)**: Captures the frequency and strength of hateful expressions in user comments. Previous research has highlighted the importance of normalizing hate prevalence using reference baselines (Waseem and Hovy, 2016; Davidson et al., 2017), but our formulation provides a media-specific perspective.

4. **Hate Polarization Index by media (HPI_m)**: Measures the proportion of extreme hate in user comments across media. This extends prior polarization indices (Agresti, 2002; Conover et al., 2011) by incorporating hate speech severity into the polarization framework.
5. **Hate Polarization Index by media topic (HPI_{mp})**: Evaluates the extent of extreme hate in user comments across media topics. While previous studies have analyzed topic-dependent hate speech variations (Agresti, 2002; Conover et al., 2011), our index provides a standardized way to compare polarization trends across different thematic categories.

The mathematical formulation of these indices is presented below.

Consider a set of media outlets indexed by i. For each outlet i, let Pos_i be the total number of comments labelled as positive, Neg_i be the total number of comments labelled as negative, and Tot_i be the total number of comments. We define the following:

$$PI_{m,i} = \frac{\text{Pos}_i - \text{Neg}_i}{\text{Pos}_i + \text{Neu}_i + \text{Neg}_i} \tag{A.9}$$

$$II_{m,i} = \frac{\text{Pos}_i + \text{Neg}_i}{\text{Pos}_i + \text{Neu}_i + \text{Neg}_i} \tag{A.10}$$

Let HateExt_i be the number of *extreme hateful* comments associated with outlet i. The *Hate Intensity* index is:

$$HII_{m,i} = \frac{\text{HateExt}_i}{\text{Tot}_i} \tag{A.11}$$

This ratio normalizes the raw count of extremely hateful comments by the total number of comments, ensuring the measure is comparable across outlets of different sizes.

To capture how much an outlet deviates from the overall tendency, let

$$\begin{aligned} p_i &= \frac{\text{HateExt}_i}{\text{Tot}_i} = HII_{m,i}, \\ p_{\text{global}} &= \frac{\sum_i \text{HateExt}_i}{\sum_i \text{Tot}_i}. \end{aligned} \tag{A.12}$$

We define the *Hate Polarization* index via a z-score:

$$HPI_{m,i} = \frac{p_i - p_{\text{global}}}{\sqrt{p_{\text{global}}\big(1 - p_{\text{global}}\big)/\text{Tot}_i}} \tag{A.13}$$

This reflects how many standard deviations above or below the global mean p_{global} the proportion p_i is, taking into account the total number of comments Tot_i for that outlet.

Appendix A.5.1. Models, preprocessing, thresholds, and validation

Models. For sentiment, we use the RoBERTa-base[1] model (Barbieri et al., 2020), pre-trained on ∼58M tweets, fine-tuned on the TweetEval sentiment task; labels: *negative*, *neutral*, *positive*.[2] We report TweetEval's task definition and metric (macro-averaged recall for sentiment) and note that RoBERTa variants trained/continued on Twitter achieve competitive scores on the unified benchmark. For hate speech, we use `pysentimiento`'s English hate-speech detector (BERTweet backbone), trained on SemEval-2019 Task 5 (HatEval) with multi-label outputs: HS (hate/no-hate), TR (targeted vs. untargeted), AG (aggressive vs. not) (Pérez et al., 2023).

Preprocessing. For sentiment, we follow the TweetEval/CardiffNLP normalization: user handles → `@user`, URLs → `http`, preserving emojis/hashtags and case; tokenization is RoBERTa BPE. This preserves negation cues (e.g., "not", "never"), slang, and OOV variants via subword segmentation and in-domain pretraining. For hate speech, we apply `pysentimiento.preprocessing`: replace handles/URLs with special tokens and shorten character elongations; emojis/hashtags and punctuation are retained. This mitigates noise while keeping polarity/negation carriers.

Inference and mapping to indices. Sentiment classifier returns class posteriors (p_{neg}, p_{neu}, p_{pos}); we assign the arg max label per comment and aggregate counts to compute PI_m and II_m (Eqs. A.9–A.10). Hate speech returns probabilities for HS, TR, and AG. We operationalize *extreme hate* as HII_m and HPI_m then use the resulting counts (Eqs. A.11–A.13).

Negation and slang handling. No rule-based negation flipping is applied; negation is modeled in-context by transformers. Slang, orthographic variants, hashtags, and emojis are handled via subword tokenization and Twitter-domain pretraining (RoBERTa-Twitter and BERTweet), which are designed for social-media vernacular.

[1] `cardiffnlp/twitter-roberta-base-sentiment`

[2] https://huggingface.co/cardiffnlp/twitter-roberta-base-sentiment

Sentiment and Hate Classifiers Validation. On English sentiment (TweetEval), the CardiffNLP twitter-RoBERTa model is evaluated with macro-averaged recall (the official metric for sentiment in TweetEval) and attains 71–73% M-Rec depending on the RoBERTa variant (the benchmark's "best" row reports 71.0%); TweetEval does not report precision or macro-F1 for this task (Barbieri et al., 2020). For pysentimiento's (pysentimiento/robertuito-sentiment-analysis), the model card reports macro-F1 = 0.705 ± 0.003 on TASS 2020; precision and recall are not provided in the card.

To provide an empirical validation of the pretrained classifiers on in-domain Italian Facebook discourse, we performed a small-scale validation on a random sample of $n = 500$ comments by manually labeling the sample. We used one annotator and reported the standard precision/recall/F1 scores. For hate, we report results under the main decision threshold $t = 0.5$ on the `hateful` score and a stricter sensitivity threshold $t = 0.7$ (Table A.7); the proxy gold prevalence is low (4/500), hence class-1 metrics are inherently unstable and should be interpreted as a conservative diagnostic rather than an independent human-labelled benchmark.

Error analysis. Inspection of misclassified cases reveals interpretable error patterns. For *sentiment*, the dominant error mode is false-positive negative (77 cases, 15.4% of the sample): the model flags as negative comments that human annotators judged neutral—typically critical but factual statements about policies, ironic remarks, or strong opinions expressed in a matter-of-fact tone. A related pattern is false-positive positive (61 cases, 12.2%), where sarcastic praise, exclamation-laden text, or dark humor with superficially positive lexicon is misread as genuine enthusiasm. False-negative errors for the negative class (22 cases, 4.4%) occur when hostility is conveyed through Italian-specific insults, regional slurs (e.g., *terrone*), or indirect phrasing that the English-pretrained model does not recognize. In sum, the primary limitation is sarcasm and irony detection: Italian political discourse frequently employs irony that the model interprets literally. For *hate speech*, the low base rate (4 hate-labelled comments, 0.8%) constrains precision. The model exhibits a high false-positive rate (137 cases at $t = 0.5$): aggressive political rhetoric, populist anti-establishment language, and emphatic Italian colloquialisms (e.g., *maledetti*, *disgraziati*) are flagged as hate speech even in the absence of group-targeted dehumanization. False negatives (2 cases, representing 50% of actual hate) arise when violence is expressed metaphorically or when hate is embedded in longer, seemingly analytical text. Given the

Table A.7: Diagnostic validation metrics (model vs. human labels) on $n = 500$ sampled comments. Sentiment is 3-class; hate is binary with thresholds on the `hateful` score.

Task	Class / Summary	Precision	Recall	F1	Support
Sentiment (3-class)					
	negative	0.306	0.627	0.411	59
	neutral	0.871	0.662	0.752	409
	positive	0.074	0.156	0.100	32
	macro avg (F1)	—	—	0.421	500
	weighted avg (F1)	—	—	0.671	500
	accuracy	—	—	0.626	500
Hate (binary), threshold $t = 0.5$					
	hate=1	0.014	0.500	0.028	4
	macro avg (F1)	—	—	0.433	500
	weighted avg (F1)	—	—	0.831	500
	accuracy	—	—	0.722	500
Hate (binary), threshold $t = 0.7$					
	hate=1	0.018	0.500	0.034	4
	macro avg (F1)	—	—	0.454	500
	weighted avg (F1)	—	—	0.866	500
	accuracy	—	—	0.776	500

domain shift from English Twitter to Italian Facebook and the extreme class imbalance, the 50% recall on actual hate comments represents reasonable—though imperfect—sensitivity; accordingly, hate-intensity indices should be interpreted as conservative upper bounds on hostility prevalence.

Appendix A.6. Validation and Robustness

We examine: (i) STM stability across seeds and $K \pm 1$; (ii) sensitivity to $\alpha \in \{0, 0.5, 1\}$ in the elastic-net; (iii) outlet/event exclusion tests; (iv) alternative comment aggregation windows. For each, we refit models and compare effect profiles via rank correlations and overlapping CIs.

Appendix A.7. Auxiliary post–comment lexical divergence in a shared space

This appendix reports an auxiliary distance-based analysis designed to address RQ2 (echo vs. expansion) in a shared representation space *without* requiring an arbitrary identification map between latent topic axes of two independently estimated STMs. Distances such as cosine similarity or Jensen–Shannon divergence between topic-mixture vectors require a common, aligned basis. With separate fits for posts and comments,

$$\theta_p^{(p)} \in \Delta^{K_{\text{post}}-1}, \quad \phi_c^{(c)} \in \Delta^{K_{\text{comm}}-1}, \quad K_{\text{post}} \neq K_{\text{comm}} \text{ (in general)}, \tag{A.14}$$

and latent axes are identifiable only up to permutation/rotation. Hence, for any distance $d(\cdot, \cdot)$ defined on matched coordinates, $d\left(\theta_p^{(p)}, \phi_c^{(c)}\right)$ is not well-defined without an identification map $\Pi : \{1, \ldots, K_{\text{post}}\} \to \{1, \ldots, K_{\text{comm}}\}$

Ad hoc mappings induce uncontrolled arbitrariness, we thus compute a thread level divergence metric comparing each post and its aggregated comments.

Definition.. Let V_{post} denote the post STM vocabulary. For a post p, let P_p be the normalized term-frequency distribution over V_{post} induced by the post text. Let Q_p be the analogous distribution induced by the *aggregate* of all comments associated with p, after projecting comment tokens onto V_{post} (dropping tokens not in V_{post}). We define the post–comment lexical divergence as the Jensen–Shannon distance

$$D_{\text{JS}}(P_p, Q_p) = \sqrt{\frac{1}{2}\text{KL}(P_p \| M_p) + \frac{1}{2}\text{KL}(Q_p \| M_p)}, \qquad M_p = \frac{1}{2}(P_p + Q_p), \tag{A.15}$$

where $\mathrm{KL}(\cdot\|\cdot)$ denotes Kullback–Leibler divergence (with standard numerical smoothing in implementation). This yields a symmetric distance in a common coordinate system.

Coverage diagnostics.. Because the distance is computed in a post-anchored space, we report the fraction of comment tokens retained after projection onto V_{post}. Overall comment-token coverage is 47.28%, with modest variation across media outlets (Table A.8).

Media outlet	n_{comments}	Total tokens	Retained tokens	Coverage
`Resto_del_Carlino`	2531	23791	10393	43.68%
`Il_Post`	5555	50391	22770	45.19%
`La_Repubblica`	28156	259600	121418	46.77%
`Avvenire`	1936	19899	9371	47.09%
`Il_Giornale`	18570	147550	69720	47.25%
`Libero`	17350	143322	67804	47.31%
`Il_Mattino`	8279	74194	35240	47.50%
`Corriere_della_Sera`	16656	151286	72401	47.86%
`Fanpage_it`	19011	155142	74263	47.87%
`Il_Fatto_Quotidiano`	21748	191162	91645	47.94%

Table A.8: Comment-token coverage in the post vocabulary space V_{post} (auxiliary shared-space divergence analysis).

Summary results. The divergence is computable for all posts. The distribution is concentrated at high values (median 0.786, IQR [0.765, 0.807], mean 0.784, min 0.682, max 0.833), indicating that comment threads are not simple lexical echoes of the initiating post within the shared post-vocabulary space. Differences across media outlets are statistically detectable but moderate in magnitude (Kruskal–Wallis $p \approx 2 \times 10^{-12}$; effect-size proxy $\varepsilon^2 \approx 0.097$). Finally, divergence exhibits a weak negative association with thread size (Spearman $\rho \approx -0.096$, $p \approx 0.012$).

Figures. Figure A.17 indicates that post–comment lexical divergence is consistently high and tightly distributed (median ≈ 0.786; IQR $\approx [0.765, 0.807]$), suggesting that comment threads are not simple lexical echoes of the initiating post within the shared post-vocabulary space.

Media outlet	n_{posts}	Median	Q1	Q3
Il_Post	79	0.7649	0.7479	0.7899
La_Repubblica	57	0.7768	0.7434	0.7947
Libero	90	0.7803	0.7690	0.7983
Fanpage_it	47	0.7851	0.7665	0.8087
Corriere_della_Sera	47	0.7852	0.7599	0.7989
Il_Fatto_Quotidiano	109	0.7871	0.7673	0.7993
Il_Giornale	118	0.7925	0.7706	0.8094
Avvenire	54	0.7934	0.7741	0.8160
Resto_del_Carlino	30	0.7959	0.7703	0.8110
Il_Mattino	58	0.8090	0.7930	0.8210

Table A.9: Auxiliary post–comment lexical divergence by media outlet (Jensen–Shannon distance; post-vocabulary space).

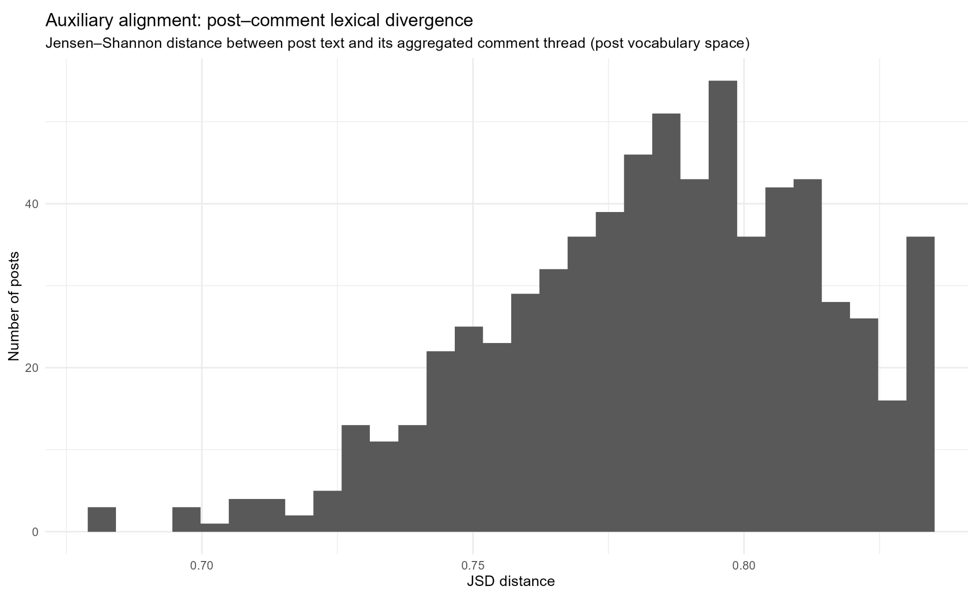

Figure A.17: Auxiliary alignment: distribution of post–comment lexical divergence. Jensen–Shannon distance between post text and aggregated comment thread in the post-vocabulary space.

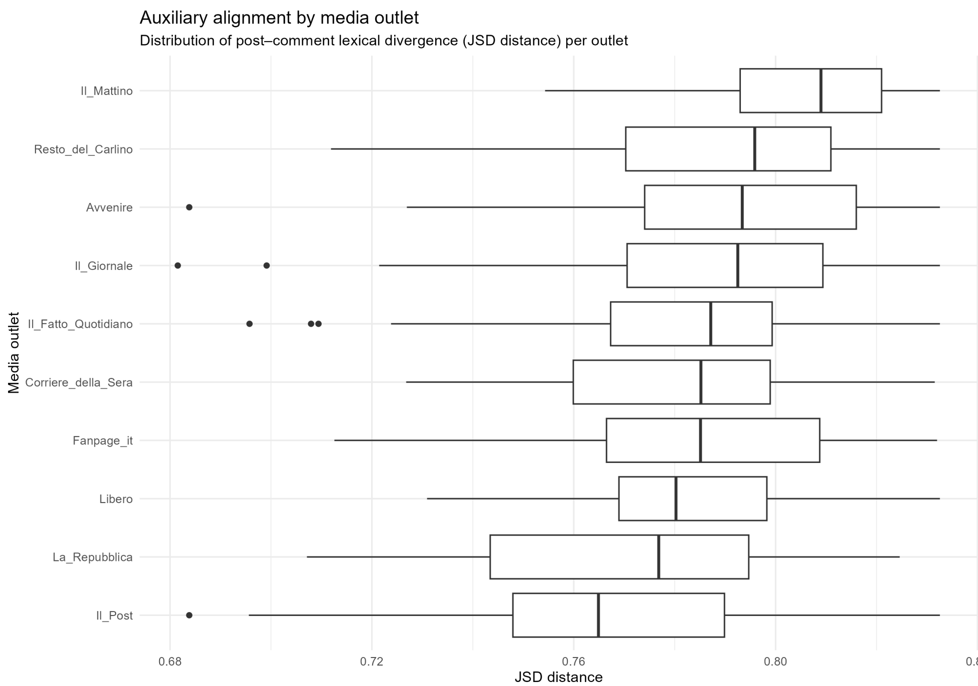

Figure A.18: Auxiliary alignment by media outlet: post–comment lexical divergence (Jensen–Shannon distance) by outlet.

Figure A.18 shows that divergence differs across outlets but with moderate between-outlet separation, consistent with outlet-specific framing effects while preserving a common pattern of expansion in user discourse.

Appendix A.8. Cluster-aware sensitivity analysis: comment-level hierarchical specification

The main analysis targets thread-level audience response to each media post and therefore operates on post-aggregated comment outcomes. As a robustness check against within-thread dependence and potential aggregation artefacts, we additionally estimated post→comment associations on individual comments using a hierarchical specification with post- and outlet-level clustering.

Let $\theta_{ik}^{(c)}$ be the estimated topic proportion for comment i on comment-topic k, and let $\theta_{j\ell}^{(p)}$ be the estimated topic proportion for post j on post-topic ℓ (from the post STM). For each comment-topic k, we fit:

$$\operatorname{logit}\left(\theta_{ik}^{(c)}\right) = \alpha_k + \sum_{\ell=1}^{K_p} \beta_{k\ell}\, \theta_{j(i)\ell}^{(p)} + u_{j(i)} + v_{o(i)} + \varepsilon_{ik}, \tag{A.16}$$

where $j(i)$ indexes the post associated with comment i, $o(i)$ indexes the outlet account (`testata`), and $u_j \sim \mathcal{N}(0, \sigma_u^2)$ and $v_o \sim \mathcal{N}(0, \sigma_v^2)$ are random intercepts capturing within-thread and within-outlet dependence. For numerical stability, the logit transformation was applied to clipped proportions. To facilitate comparison across topic pairs, we report standardized coefficients $\beta_{k\ell}^{*} = \beta_{k\ell}\, \mathrm{sd}(\theta_\ell^{(p)})/\mathrm{sd}(\operatorname{logit}(\theta_k^{(c)}))$.

Figure A.19 visualizes the full grid of standardized effects $\{\beta_{k\ell}^{*}\}$ as a heatmap. The estimated associations are concentrated on a limited subset of topic pairs and remain small in magnitude overall (Table A.10), consistent with the interpretation of the post→comment links as weak-to-moderate shifts in thread composition rather than strong deterministic mappings.

To align this sensitivity analysis with the thread-level estimand, we further aggregated comment-level fitted values back to the post level and compared predicted vs. observed per-post mean comment-topic shares. Table A.11 reports topic-wise correlations (range 0.258–0.653, median 0.506), indicating non-trivial agreement between hierarchical comment-level predictions and observed post-level averages.

Summary statistic (over all (k, ℓ) pairs)	Value
$\max_{k,\ell} \lvert\beta^*_{k\ell}\rvert$	0.0776
$\text{mean}_{k,\ell} \lvert\beta^*_{k\ell}\rvert$	0.0177
$\Pr(\lvert\beta^*_{k\ell}\rvert > 0.01)$	0.600
$\Pr(\lvert\beta^*_{k\ell}\rvert > 0.02)$	0.400
$\Pr(\lvert\beta^*_{k\ell}\rvert > 0.03)$	0.133

Comment topic	Post topic	β	SE	β^*
C7	P4	-0.1053	0.0253	-0.0776
C7	P3	-0.0707	0.0278	-0.0395
C7	P2	-0.0649	0.0266	-0.0392
C7	P1	-0.0525	0.0223	-0.0391
C12	P4	0.0579	0.0292	0.0373
C13	P3	0.0660	0.0305	0.0362
C8	P1	-0.0567	0.0335	-0.0324
C15	P4	0.0455	0.0295	0.0307
C5	P1	0.0440	0.0278	0.0287
C12	P2	0.0540	0.0305	0.0285

Table A.10: Comment-level hierarchical sensitivity: compact summary of standardized coefficients β^* from Eq. (A.16) and the strongest topic-pair associations.

RTOM

Comment topic	Correlation
C1	0.3598
C2	0.5455
C3	0.5575
C4	0.6533
C5	0.4310
C6	0.4196
C7	0.2583
C8	0.5615
C9	0.6170
C10	0.4197
C11	0.4687
C12	0.6526
C13	0.5590
C14	0.5060
C15	0.4905

Table A.11: Agreement between observed per-post mean comment-topic shares and hierarchical-model predictions aggregated at the post level (topic-wise correlations).

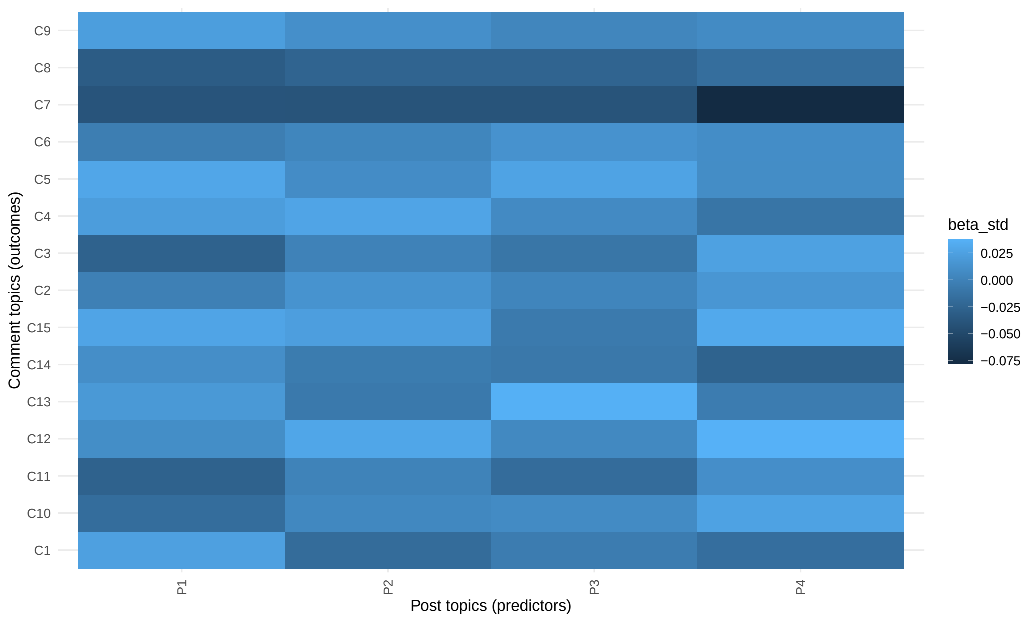

Figure A.19: Comment-level hierarchical sensitivity: heatmap of standardized coefficients β^* from Eq. (A.16).

References

AGCOM (2024a). Communications observatory no. 3/2024 (english edition). Technical report, Autorità per le Garanzie nelle Comunicazioni (AGCOM), Rome, Italy. Social networking services: monthly unique users and time spent.

AGCOM (2024b). Osservatorio sulle comunicazioni n. 2/2024. Technical report, Autorità per le Garanzie nelle Comunicazioni (AGCOM), Rome, Italy. Section on social network unique users and audience metrics.

Agresti, A. (2002). *Categorical Data Analysis*. Wiley-Interscience.

Barbieri, F., Camacho-Collados, J., Espinosa Anke, L., and Neves, L. (2020). TweetEval: Unified benchmark and comparative evaluation for tweet classification. In Cohn, T., He, Y., and Liu, Y., editors, *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 1644–1650, Online. Association for Computational Linguistics.

Bartlett, J. and Norrie, R. (2015). Immigration on twitter: understanding public attitudes online. *Demos*, 1:1–48.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., and Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, 5(4):323–370.

Burnap, P. and Williams, M. L. (2015). Cyber hate speech on twitter: An application of machine classification and statistical modeling for policy and decision making. *Policy & internet*, 7(2):223–242.

Burnap, P. and Williams, M. L. (2016). Us and them: identifying cyber hate on twitter across multiple protected characteristics. *EPJ Data science*, 5(1):11.

Calderón, C. A., de la Vega, G., and Herrero, D. B. (2020). Topic modeling and characterization of hate speech against immigrants on twitter around the emergence of a far-right party in spain. *Social Sciences*, 9(11):188.

Capozzi, A., de Francisci Morales, G., Mejova, Y., Monti, C., Panisson, A., and Paolotti, D. (2021). Clandestino or rifugiato? anti-immigration facebook ad targeting in italy. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–15.

Cohen, S. (1972). *Folk Devils and Moral Panics: The Creation of the Mods and Rockers*. MacGibbon and Kee, London.

Conover, M. D., Ferrara, E., Menczer, F., and Flammini, A. (2011). Political polarization on twitter. In *Proceedings of the International AAAI Conference on Weblogs and Social Media*, pages 89–96.

Davidson, T., Warmsley, D., Macy, M., and Weber, I. (2017). Automated hate speech detection and the problem of offensive language. In *Proceedings of the International AAAI Conference on Web and Social Media (ICWSM)*.

Drouhot, L. G., Deutschmann, E., Zuccotti, C. V., and Zagheni, E. (2023). Computational approaches to migration and integration research: promises and challenges. *Journal of Ethnic and Migration Studies*, 49(2):389–407.

Entman, R. M. (1993). Framing: Towards clarification of a fractured paradigm. *McQuail's reader in mass communication theory*, 390:397.

Flores, R. D. (2017). Do anti-immigrant laws shape public sentiment? a study of arizona's sb 1070 using twitter data. *American Journal of Sociology*, 123(2):333–384.

Fortuna, P. and Nunes, S. (2018). A survey on automatic detection of hate speech in text. *Acm Computing Surveys (Csur)*, 51(4):1–30.

Garland, J., Ghazi-Zahedi, K., Young, J.-G., Hébert-Dufresne, L., and Galesic, M. (2022). Impact and dynamics of hate and counter speech online. *EPJ data science*, 11(1):3.

Grigorieff, A., Roth, C., and Ubfal, D. (2020). Does information change attitudes toward immigrants? *Demography*, 57:1117–1143.

Gualda, E. and Rebollo, C. (2016). The refugee crisis on twitter: A diversity of discourses at a european crossroads. *Journal of Spatial and Organizational Dynamics*, 4(3):199–212.

Helbling, M. (2014). Framing immigration in western europe. *Journal of Ethnic and Migration Studies*, 40(1):21–41.

Hovland, C. I., Janis, I. L., and Kelley, H. H. (1953). Communication and persuasion.

Hurtado Bodell, M., Magnusson, M., and Keuschnigg, M. (2024). Seeded topic models in digital archives: Analyzing interpretations of immigration in swedish newspapers, 1945–2019. *Sociological Methods & Research*, page 00491241241268453.

Iyengar, S. (1994). *Is anyone responsible?: How television frames political issues*. University of Chicago Press.

Iyengar, S. and Kinder, D. R. (2010). *News that matters: Television & American opinion*. University of Chicago Press.

Iyengar, S. and Kinder, D. R. (2024). *News that matters: Television & American opinion*. University of Chicago Press.

Kasperson, R. E., Renn, O., Slovic, P., Brown, H. S., Emel, J., Goble, R., Kasperson, J. X., and Ratick, S. (1988). The social amplification of risk: A conceptual framework. *Risk Analysis*, 8(2):177–187.

Katz, E., Lazarsfeld, P. F., and Roper, E. (2017). *Personal influence: The part played by people in the flow of mass communications*. Routledge.

Kelling, C. and Monroe, B. L. (2023). Analysing community reaction to refugees through text analysis of social media data. *Journal of Ethnic and Migration Studies*, 49(2):492–534.

Kim, J., Zhang, Y., and Borah, P. (2025). Second-level agenda-setting effects of news media and public policy on social media discourse across platforms: immigration during the covid-19 pandemic in the us. *Information, Communication & Society*, 28(11):1923–1944.

Kriesi, H. (2012). *Political conflict in western Europe*. Cambridge University Press.

Liu, B. (2012). *Sentiment Analysis and Opinion Mining*, volume 5 of *Synthesis Lectures on Human Language Technologies*. Morgan & Claypool Publishers.

Lord, C. G., Ross, L., and Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, 37(11):2098–2109.

Mazzoleni, G. and Bracciale, R. (2018). Socially mediated populism: the communicative strategies of political leaders on facebook. *Palgrave Communications*, 4(1).

McCombs, M. E. and Shaw, D. L. (1972). The agenda-setting function of mass media. *Public Opinion Quarterly*, 36(2):176–187.

Mehrabian, A. and Russell, J. A. (1974). *An Approach to Environmental Psychology*. MIT Press, Cambridge, MA.

Menshikova, A. and van Tubergen, F. (2022). What drives anti-immigrant sentiments online? a novel approach using twitter. *European Sociological Review*, 38(5):694–706.

Merrill, S. and Åkerlund, M. (2018). Standing up for sweden? the racist discourses, architectures and affordances of an anti-immigration facebook group. *Journal of Computer-Mediated Communication*, 23(6):332–353.

Miccoli, S., Ambrosetti, E., et al. (2023). The perceptions of migration during the pandemic: what twitter data tell us. *Migration Letters*, 20(2):189–198.

Monroe, B. L., Colaresi, M. P., and Quinn, K. M. (2008). Fightin'words: Lexical feature selection and evaluation for identifying the content of political conflict. *Political Analysis*, 16(4):372–403.

Pang, B. and Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2):1–135.

Pierri, F. (2024). Drivers of hate speech in political conversations on twitter: the case of the 2022 italian general election. *EPJ Data Science*, 13(1):63.

Prislei, L., Di Cicco, G., Sensales, G., and d'Urso, E. (2022). Representations of immigration on facebook: A lexicographic analysis of the communication of italian populist and non-populist leaders (2019-2021). *Psychology Hub*, 39(3):41–52.

Pérez, J. M., Rajngewerc, M., Giudici, J. C., Furman, D. A., Luque, F., Alemany, L. A., and Martínez, M. V. (2023). pysentimiento: A python toolkit for opinion mining and social nlp tasks.

Reuters (2024). Digital news report 2024. Global and country findings on news use via platforms.

Reuters (2025). Digital news report 2025. Methodology and country-level usage of social platforms for news.

Righi, A. (2019). Assessing migration through social media: a review. *Mathematical Population Studies*, 26(2):80–91.

Robert, D. and John, R. (1982). Store atmosphere: an environmental psychology approach. *Journal of retailing*, 58(1):34–57.

Rowe, F., Mahony, M., Graells-Garrido, E., Rango, M., and Sievers, N. (2022). Using twitter to track immigration sentiment during early stages of the covid-19 pandemic—addendum. *Data & Policy*, 4:e11.

Scheufele, D. A. and Tewksbury, D. (2007). Framing, agenda setting, and priming: The evolution of three media effects models. *Journal of communication*, 57(1):9–20.

Shaw, D. L., McCombs, M., Weaver, D. H., and Hamm, B. J. (1999). Individuals, groups, and agenda melding: A theory of social dissonance. *International Journal of Public Opinion Research*, 11(1):2–24.

Soroka, S. N. (2014). *Negativity in Democratic Politics: Causes and Consequences*. Cambridge Studies in Public Opinion and Political Psychology. Cambridge University Press, Cambridge.

Stephan, W. G. and Stephan, C. W. (2000). An integrated threat theory of prejudice. In Oskamp, S., editor, *Reducing Prejudice and Discrimination*, pages 23–46. Lawrence Erlbaum Associates, Mahwah, NJ.

Stroud, N. J. (2011). *Niche News: The Politics of News Choice*. Oxford University Press, New York.

Tetlock, P. C., Saar-Tsechansky, M., and Macskassy, S. (2008). More than words: Quantifying language to measure firms' fundamentals. *The Journal of Finance*, 63(3):1437–1467.

Vallone, R. P., Ross, L., and Lepper, M. R. (1985). The hostile media phenomenon: Biased perception and perceptions of media bias in coverage of the beirut massacre. *Journal of Personality and Social Psychology*, 49(3):577–585.

Viola, L. and Verheul, J. (2020). Mining ethnicity: Discourse-driven topic modelling of immigrant discourses in the usa, 1898–1920. *Digital Scholarship in the Humanities*, 35(4):921–943.

Waseem, Z. and Hovy, D. (2016). Hateful symbols or hateful people? predictive features for hate speech detection on twitter. In *Proceedings of the NAACL Student Research Workshop*, pages 88–93.

Wilson, T., Wiebe, J., and Hoffmann, P. (2005). Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 347–354.

Yarchi, M., Baden, C., and Kligler-Vilenchik, N. (2024). Political polarization on the digital sphere: A cross-platform, over-time analysis of interactional, positional, and affective polarization on social media. In *Dissonant Public Spheres*, pages 185–226. Routledge.

Zhang, Y. and Lian, R. (2025). How right-wing media, twitter, facebook, and youtube circulated antipathy and threat cues about immigrants: A cross-media and cross-platform approach. *Mass Communication and Society*, pages 1–20.