



# Navigating the brain drain crisis: Knowledge, visibility, sentiment, and divergence on X (formerly Twitter) among English, Spanish, and Italian users

Jimi Park<sup>a,\*</sup>, Nicole Pitassi<sup>b</sup>

<sup>a</sup> Hult International Business School, 1 Education Street, Cambridge, MA 0214, USA

<sup>b</sup> University College London, Department of Political Science, 29, 30 Tavistock Square, London WC1H 9QU, United Kingdom

## ARTICLE INFO

### Keywords:

X (formerly twitter) discourse  
Sentiment shift  
Topic divergence  
Social capital  
Migration discourse  
F42 international policy coordination and transmission  
O35 social innovation  
M38 marketing and advertising: Government policy and regulation

## ABSTRACT

Brain drain, the emigration of highly skilled workers, is both a pressing policy issue and a cultural frame for examining how societies negotiate belonging, loss, and opportunity. Using data from X (formerly known as Twitter) across English, Spanish, and Italian debates, this study examines engagement dynamics through three dimensions: visibility, sentiment shifts, and thematic divergence. In the present context, we define visibility as the weighted exposure of replies and likes that are standardized within each language. The dataset includes approximately 3100 English, 4500 Spanish, and 250 Italian tweet-reply pairs collected between 2020 and 2024. The findings show that visibility consistently drives engagement but carries distinct cultural meanings. The discourse in English illustrates coherence: engagement is amplified when replies remain thematically aligned, reinforcing bonding capital and collective narratives. The Spanish discourse illustrates plurality: visible divergent replies broaden participation, reflecting bridging capital while introducing volatility. The Italian discourse illustrates optimism: positive sentiment shifts sustain institutional trust with divergence. For policymakers, these results highlight the following specific pathways to tailor crisis communication: reinforce coherence in English debates, embrace plurality in Spanish contexts, and sustain optimism in Italian discourse.

## Introduction

Brain drain, the emigration of highly skilled workers, remains both a critical policy challenge and a lens that reveals how societies negotiate opportunity, belonging, and loss. In an era when debates on migration increasingly unfold on digital platforms such as X (formerly Twitter), social media functions not merely as a communication tool but as a contested arena where narratives are shaped, amplified, and resisted. Understanding these dynamics requires moving beyond descriptive accounts of online discourse to frameworks that capture the scale and nature of the structural, affective, and thematic patterns at play.

Language plays a pivotal role in this process. It functions simultaneously as a structural filter that shapes the participants of digital public spheres and a cultural framing device that conditions how migration narratives are articulated and received (Benson, 2016; Dubois & Blank, 2018; Hale et al., 2022; Estevez et al., 2025). Linguistic segmentation is therefore more than a methodological choice: it is a theoretical necessity for understanding how publics process crises through distinct cultural

logics. The examination of English, Spanish, and Italian brain drain debates highlights three contrasting contexts: post-Brexit labor uncertainty, Latin American migration pressures, and Italian institutional inertia, each embedded within different linguistic-cultural frames.

To analyze these dynamics, this study integrates social capital theory with computational methods. Social capital distinguishes between bonding capital, which reinforces cohesion through thematic alignment and shared sentiment, and bridging capital, which expands networks by incorporating divergent perspectives (Coleman, 1988; Putnam, 2000). Retweeting is not merely amplification but a relational signal of trust, reciprocity, or contestation within linguistic communities. We operationalize three pillars of retweet behavior to capture these dynamics: visibility, measuring the structural reach of tweets and replies; sentiment shift, capturing the affective modulation between tweets and replies; and divergence, which indicates whether conversations remain thematically coherent or diverge into new directions. Through regression modeling and divergence metrics, computational approaches allow us to move beyond anecdotal contrasts and systematically reveal how

\* Corresponding author.

E-mail addresses: [jimi.park@faculty.hult.edu](mailto:jimi.park@faculty.hult.edu) (J. Park), [uctqnpi@ucl.ac.uk](mailto:uctqnpi@ucl.ac.uk) (N. Pitassi).

<https://doi.org/10.1016/j.jik.2025.100906>

Received 3 July 2025; Accepted 28 November 2025

Available online 4 December 2025

2444-569X/© 2025 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

publics process migration debates in different languages.

The results show that visibility consistently drives engagement but operates differently across linguistic communities. English retweet dynamics penalize divergence while privileging thematic coherence and emotional restraint, which is consistent with bonding capital. By contrast, Spanish dynamics selectively reward visible divergent replies, reflecting bridging capital and the cultural value of expressive plurality. Italian dynamics place less weight on divergence and amplify engagement through positive sentiment shifts, which sustain institutional trust and optimism. Interaction effects across models underscore how structural signals (visibility), affective modulations (sentiment), and thematic alignment (divergence) are interpreted through distinct cultural filters.

These findings contribute to the theory by demonstrating how language structures both participation and the legitimacy of expression in the discourse on migration, linking digital publics to broader debates on transnational communication. The implications for crisis management are clear. English publics consolidate around coherent, professionalized narratives that support coordinated messaging and risk homogenization. Spanish publics thrive on plurality, where the selective amplification of divergent voices enhances inclusivity while raising the costs of emotional labor. Italian publics sustain engagement through institutional optimism and highlight the importance of affective framing in maintaining trust. Together, these patterns show that effective crisis communication cannot rely on a one-size-fits-all strategy: it must adapt to the cultural logics embedded in language, leveraging visibility, sentiment, and divergence differently to sustain credibility and engagement in multilingual public spheres. To anchor this analysis further, Table 1 introduces a three-level categorization—individual, collective, and organizational—that defines how the discourse on brain drain manifests across linguistic publics, with multilingual examples illustrating the spectrum from personal experience to societal framing.

## Literature review

### *Migration narratives and digital public spheres*

Research on migration has long emphasized that mobility is not merely an economic or demographic process but also a communicative one (Hynek et al., 2025). Narratives of departure, opportunity, and loss are socially constructed and politically contested across multiple arenas (Docquier & Rapoport, 2012; Benson, 2016; Porfirio et al., 2024). With the rise of social media, the discourse on migration has increasingly unfolded within the digital public sphere, where affect, visibility, and participation are shaped by platform affordances (Dubois & Blank, 2018; Hale et al., 2022). X, in particular, functions as a transnational arena where narratives of “brain drain” circulate, interact, and compete for legitimacy across cultural and linguistic contexts.

### *Linguistic segmentation as a structural filter and cultural frame*

Language is more than a neutral medium; it structures participation and frames interpretation. Linguistic boundaries act as a structural filter, defining who can participate in discourse communities and shaping the flow of information across them. Language also serves as a cultural framing device, influencing how narratives are articulated, the affective registers deemed legitimate, and the forms of authenticity that are rewarded (Benson, 2016; Dubois & Blank, 2018). Comparative work in cross-linguistic social media analysis<sup>1</sup> shows that English, Spanish, and Italian discourses often emphasize policy coherence, expressive plurality, and institutional trust and optimism, respectively (Ibáñez-Sánchez et al., 2022; Kim et al., 2025). Without segmentation by language, these cultural logics are flattened, obscuring the core dynamics that shape

migration narratives in global digital arenas.

### *Social capital and the three pillars of retweet behavior*

To interpret cross-linguistic engagement patterns, this study draws on social capital theory (Bourdieu, 1986; Coleman, 1988; Putnam, 2000). Social capital distinguishes between bonding capital, which reinforces cohesion and trust within communities, and bridging capital, which expands networks by incorporating divergent perspectives. Retweeting is not only an act of amplification but also a relational signal of endorsement, contestation, or belonging that is embedded in cultural norms.

We operationalize this framework through three pillars of retweet behavior. Visibility reflects the structural distribution of attention and shows how tweets and replies extend their reach across different linguistic publics. Sentiment shift captures the affective dimension of engagement and measures how the emotional tone changes between original posts and replies. Divergence assesses whether conversations maintain thematic coherence (bonding capital) or depart into new directions (bridging capital). By embedding these pillars within the social capital lens, we link computational measures of visibility, affect, and thematic distance with broader cultural logics of trust, reciprocity, and authenticity.

### *Descriptive insights on multilingual retweet patterns*

The descriptive analysis provides the empirical backbone for this study. Fig. 1a presents the most frequent terms by category (collective, organizational, individual), while Fig. 1b maps seven Latent Dirichlet Allocation (LDA) topics that structure cross-language debates. Table 2 quantifies engagement at the category level (tweets, replies, likes, sentiment, visibility, divergence), and Fig. 2 displays sentiment distributions across tweets and replies. Figs. 3a–c compare visibility, sentiment shifts, and divergence by category and topic for each language, highlighting whether publics concentrate their engagement on original tweets versus replies, whether thematic divergence is suppressed or amplified, and whether sentiment shifts stabilize or destabilize the discourse. These descriptives lead to the formation of the research questions, which are subsequently tested through regression models.

### *English: coherence through tweet visibility*

Engagement in English is concentrated in collective topics (21 tweets; 2041 replies), which also attract the highest likes per tweet (1758; Table 2). Fig. 1a highlights terms such as “UK,” “brain,” and “people,” while Fig. 1b shows causes and patriotism as dominant themes. Fig. 2 indicates neutral-to-positive replies, and Fig. 3a confirms strong tweet visibility but weaker reply visibility, minimal sentiment shift, and low divergence. These patterns suggest a logic of coherence, where engagement clusters around visible, thematically aligned content.

*RQ1 (English – Coherence): To what extent do tweet visibility reinforce coherent, collective engagement in English debates?*

### *Spanish: plurality via reply visibility and divergence*

The debates in Spanish have the highest organizational volume (93 tweets; 160 replies) and also show that individual topics generate the most traction (616 likes per tweet; Table 2). Fig. 1a presents institutional and personal terms such as *fuga* and *trabajo*, while Fig. 1b emphasizes personal narratives and economy-related themes. Fig. 2 reveals broader, often more negative sentiment distributions, and Fig. 3b shows strong reply visibility coupled with higher divergence and more pronounced sentiment shifts. This indicates a logic of plurality, where diverse voices gain traction through reply visibility; although volatility emerges when negative sentiment spreads.

<sup>1</sup> We thank the anonymous reviewer for this valuable suggestion.

**Table 1**

Category definitions with multilingual examples.

Categories	Definition	Examples
Individual level	Tweets and comments in which the user expresses personal ideas, often drawing on personal experiences or events that concern only them.	<p>English</p> <p>"I visited earthquake-affected areas in #Paktika where I witnessed the inspiring work of female healthcare workers under tremendously difficult circumstances. This comes at a time when the #Afghan health sector is struggling with cuts in funding and braindrain. #Afghanistan <a href="https://t.co/Oyjwh3qCxd">https://t.co/Oyjwh3qCxd</a>"</p> <p>Spanish</p> <p>"Leio no Estadão que a "fuga de cérebros" do Brasil para os Estados Unidos é a maior em dez anos e me pergunto: por que será que nenhum desses "cérebros" fugiu para a Venezuela?"</p> <p>Italian</p> <p>"Il prossimo che mi parla di fuga di cervelli da sud a nord dovrà chiamare un esorcista per placarmi. #irresponsabili"</p>
Collective level	Tweets and comments in which the user shares ideas and examples that involve other individuals. This level may also include responses or mentions directed at others.	<p>English</p> <p>"The brain drain down south is no joke... I find it ironic that the most uneducated people are the one's who want to reopen schools..."</p> <p>Spanish</p> <p>"@FabianWaldman Gracias por contestar Fabian. Entiendo que estés en contra de este gobierno pero yo no me enteré que haya cerrado alguna universidad... Si mal no recuerdo no fue Massa el que recortó un montón de guita a las universidades? Ahí no hubo fuga de cerebros???"</p> <p>Italian</p> <p>"#PrimoMaggio senza cortei, senza concerti in piazza e soprattutto, per troppi, senza lavoro. L'emergenza #covid_19Italia ha insegnato quanto siamo stati stolti a provocare la fuga di cervelli e talenti spesso per mandare avanti incapaci raccomandati #FestadeiLavoratori #1maggio"</p>
Organizational level	Tweets and comments in which the user conveys an idea or explains a concept that is part of broader societal contexts.	<p>English</p> <p>"Israeli universities are offering spots to people in US universities who don't feel safe. This is circulating at @MIT right now. People will certainly take them up on the offer. There is a Jewish brain drain going on elite US universities. We should all be concerned about the <a href="https://t.co/lrTBvVlfeG">https://t.co/lrTBvVlfeG</a>"</p> <p>Spanish</p> <p>"En 2021, el gobierno se sacó de la manga un programa para repatriar a los "cerebros" emigrados y contratarlos de nuevo en las Universidades españolas. Todo se pagaba con fondos europeos (a 4000€/m brutos por investigador). Ahora vence el plazo y no tienen plaza. <a href="https://t.co/dauOthJuLN">https://t.co/dauOthJuLN</a>"</p> <p>Italian</p> <p>"(2/6) La cosiddetta "fuga di cervelli" è un fenomeno tipico della società</p>

**Table 1 (continued)**

Categories	Definition	Examples
		<p>globalizzata, ma che sempre più si sta trasformando in un problema nel nostro paese e anche in #EmiliaRomagna: annualmente, in ben 3000 ragazzi e ragazze puntano all'estero."</p>

RQ2 (Spanish – Plurality): How do reply visibility and thematic divergence support plurality in Spanish debates, and how is this affected by shifts in sentiment?

Italian: optimism through positive sentiment shifts

Italian engagement is centered around organizational topics (41 tweets; 38 replies; Table 2), with consistently positive sentiment (Fig. 2). Fig. 1a highlights *lavoro* and *fuga*, while Fig. 1b anchors debate in government, education, and opportunities. Fig. 3c shows modest tweet visibility, steady reply visibility, consistently positive sentiment shifts, and low divergence. This reflects a logic of optimism, where a positive tone reinforces organizational discourse even in the absence of high visibility.

RQ3 (Italian – Optimism): How do positive sentiment shifts sustain organizational optimism in Italian debates, even when overall visibility is lower?

#### Cross-linguistic integration

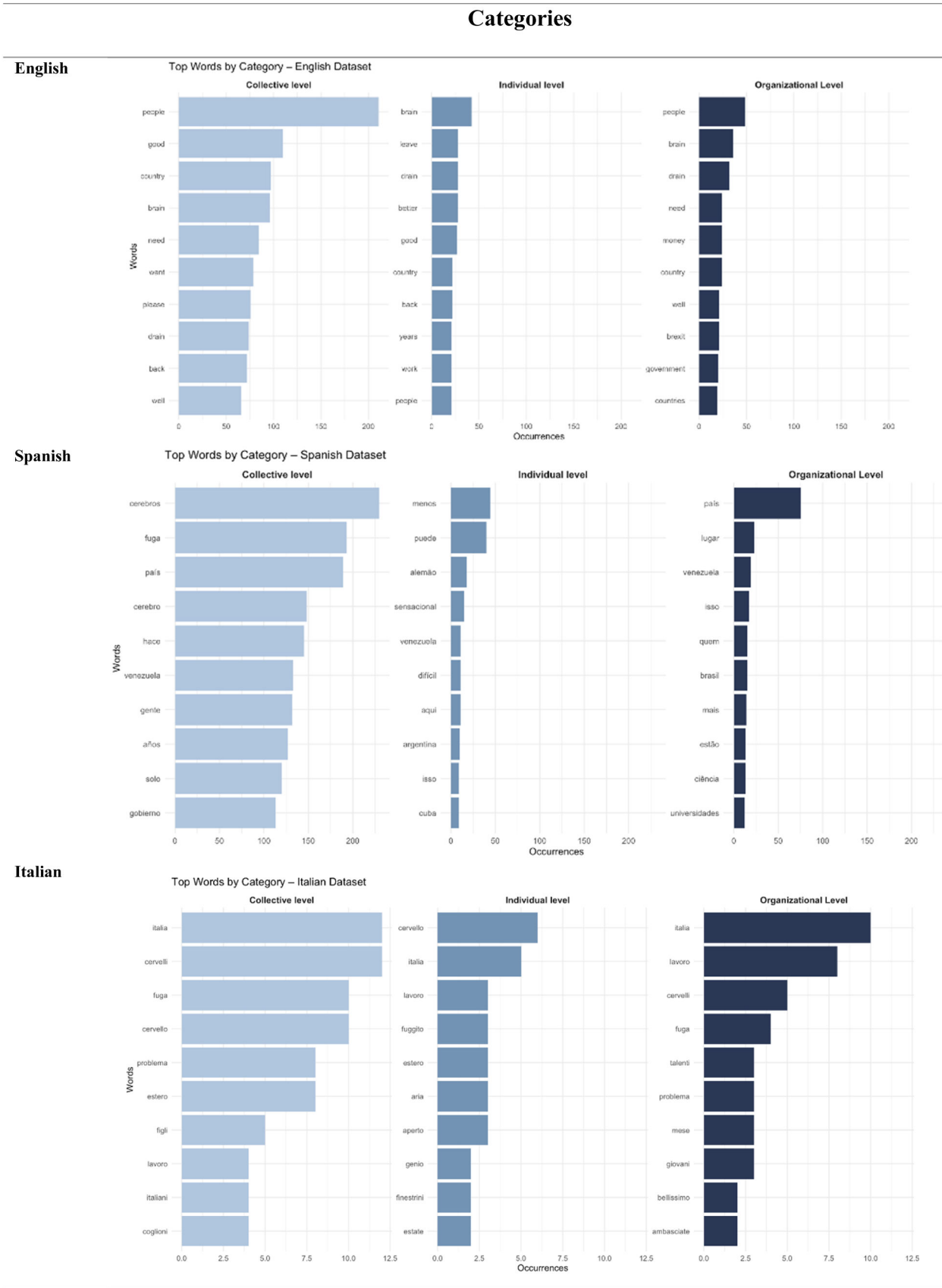
Taken together, Figs. 1 and 2 and Table 2 reveal three distinct logics: English publics rely on tweet visibility to preserve coherence, Spanish publics leverage reply visibility and divergence to amplify plurality, and Italian publics sustain optimism through positive sentiment shifts despite modest visibility. These descriptive insights establish the foundation for the regression analysis, where the interaction of visibility, sentiment, and divergence is tested to produce differentiated outcomes across linguistic publics.

#### Data and methods

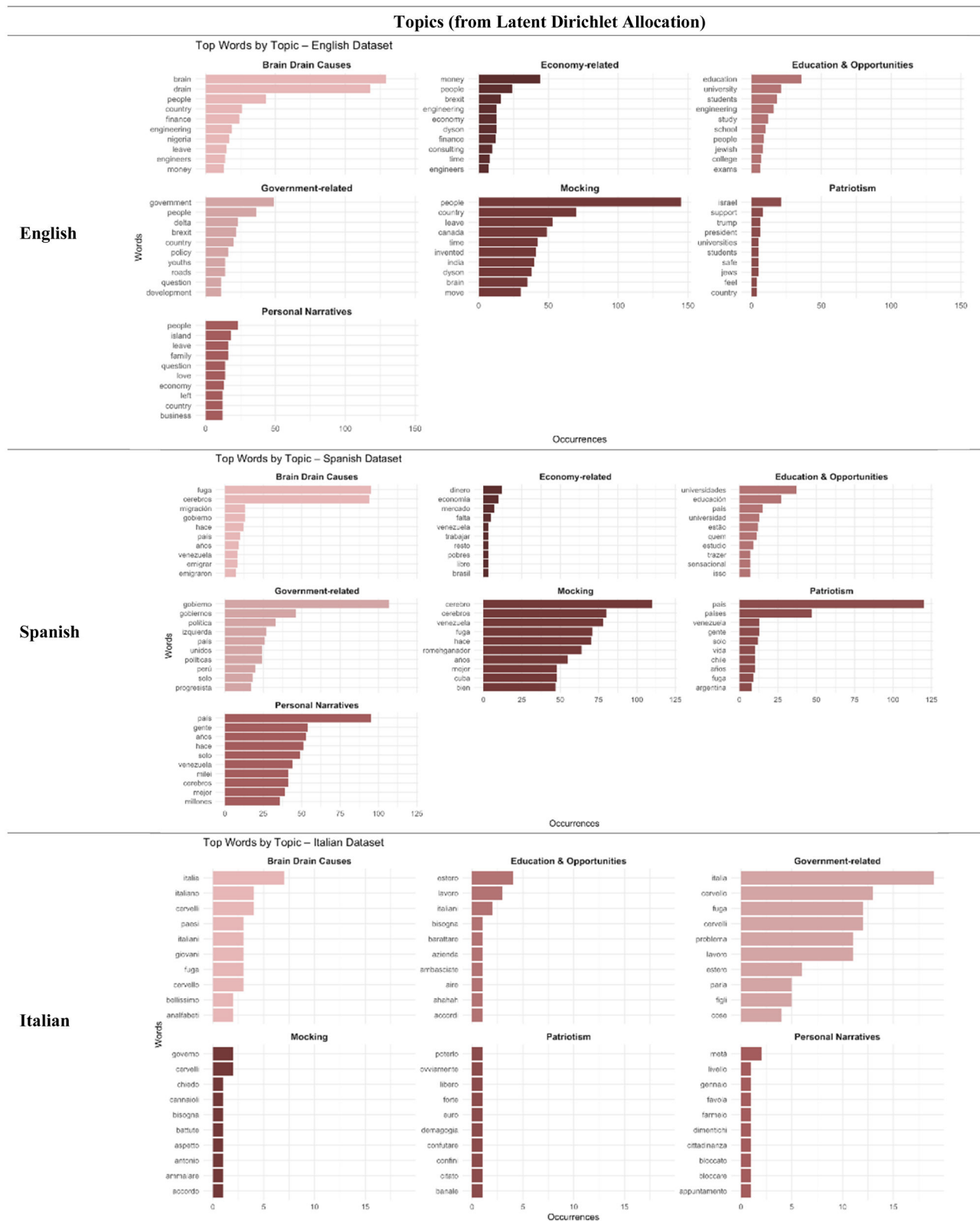
##### Data collection

We collected multilingual X data on the brain drain discourse in English, Spanish, and Italian between January 2020 and December 2024. This timeframe was chosen because migration debates intensified during this period due to post-Brexit concerns about labor, the Covid-19 pandemic, and the European energy crisis, all of which heightened anxieties over the emigration of skilled professionals (Segev, 2023; Cau et al., 2024). Simultaneously, changes to X's algorithm reshaped the patterns of visibility and engagement and rendered this window especially relevant for studying digital discourse (Brady et al., 2021; Rathje et al., 2021).

Data were retrieved using the Apify API with language-specific keywords and hashtags: "brain drain" and #braindrain for English, "fuga de cerebros" and #fugadecerebros for Spanish, and "fuga di cervelli" and #fugadicervelli for Italian. Multilingual or code-switched tweets were included if the focal language was dominant and contained the relevant term; otherwise, they were excluded. After filtering off-topic content, the dataset comprised 79 English tweets (3191 replies), 122 Italian tweets (390 replies), and 181 Spanish tweets (4508 replies). Mixed-language or off-topic posts, primarily those referencing unrelated economic or political debates, were excluded, accounting for approximately 7 % of the total corpus. The final dataset includes 3132 English, 4490 Spanish, and 253 Italian tweet-reply pairs, reflecting expected activity asymmetries across linguistic communities. The



**Fig. 1a.** Top words within categories across languages. Panels show the most frequent words per category (Individual, Collective, Organizational) for English, Spanish, and Italian tweets.



**Fig. 1b.** Latent Dirichlet Allocation (LDA) topic distributions across languages.

Note. Panels (a–c) display the top 10 representative words for each topic derived from Latent Dirichlet Allocation, grouped by language.

smaller Italian subset led to thinner cells for some interaction terms, which we interpret with caution.

Following best practices in multilingual computational social science (Charquero-Ballester et al., 2024), metadata (likes, retweets, replies, timestamps, author IDs) were preserved for traceability. A de-identified

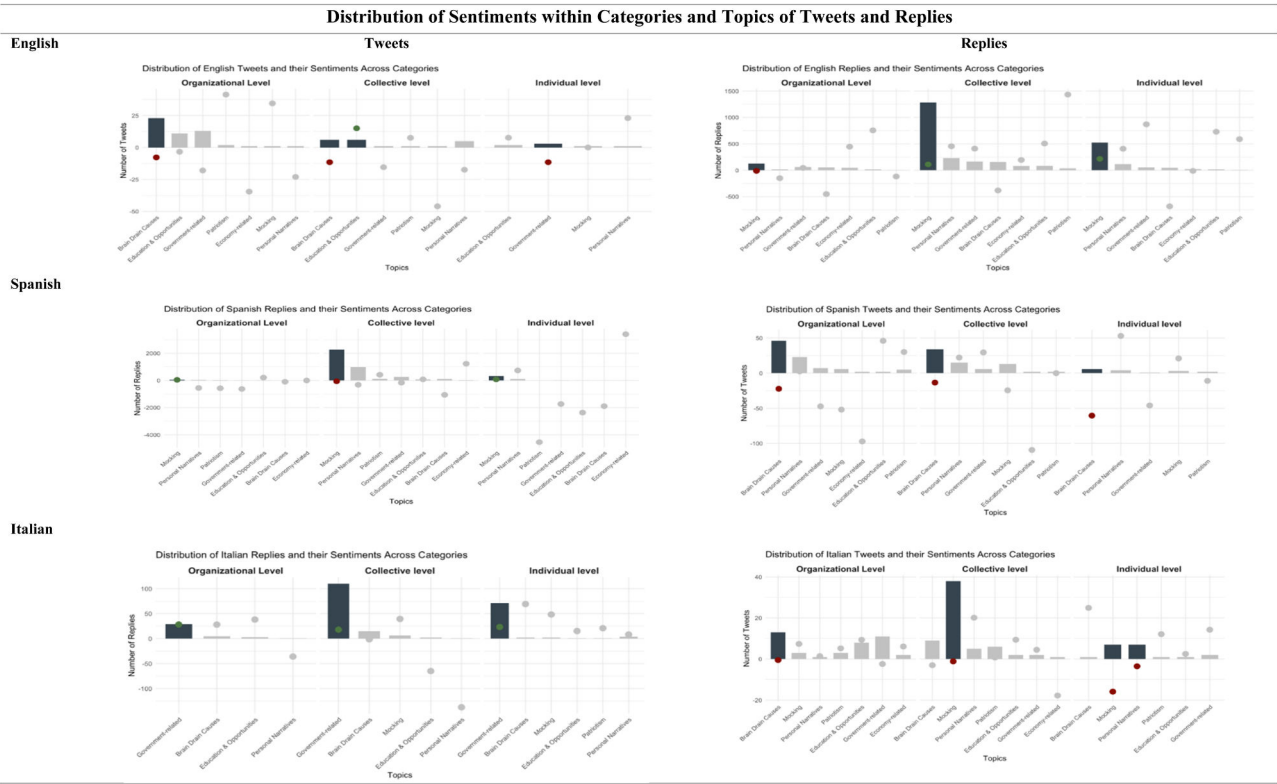
replication package containing the code and minimal data samples used in this study is publicly available through the Open Science Framework (OSF) (<https://doi.org/10.17605/OSF.IO/S6VZ2>). Manual topic validation was conducted by a single coder to ensure thematic accuracy in the tweet-reply pair classification. To verify reliability, a 10 % subset



**Table 2**  
Distribution of tweets and replies by category across languages.

Categories	# of Tweets	Average Likes for Tweets	Average Replies for Tweets	Sentiment Score Tweets	# of Replies	Average Likes for Replies	Average Sentiment Score Replies	Average Sentiment Shift	Average Tweet Visibility	Average Topic Divergence
English										
Individual	7	294	39	0.5	778	6	0.56	1.46	1352.7	0.95
Collective	21	1758	149	0.46	2041	28	0.58	1.4	175.35	0.89
Organizational	51	1553	49	0.45	328	29	0.6	1.82	242.73	0.91
Spanish										
Individual	16	616	44	0.53	251	4	0.5	1.10	102.3	0.96
Collective	72	423	28	0.5	344	11	0.5	1.32	146.49	0.89
Organizational	93	722	43	0.55	160	14	0.49	1.03	33.35	1
Italian										
Individual	19	137	6	0.63	81	1	0.52	0.48	2.88	0.83
Collective	62	11	2	0.62	134	1	0.5	0.82	7.3	0.89
Organizational	41	15	2	0.62	38	2	0.53	0.55	1.95	0.95

Note. Metrics include average likes, average replies, sentiment scores, sentiment shifts, tweet visibility, and topic divergence. Visibility measures are standardized within language.



**Fig. 2.** Sentiment distributions by category and topic.  
Note. Distribution of tweet and reply sentiment within categories and topics for each language. All sentiment scores are standardized within language (z-scores).

(approximately 500 tweet-reply pairs, proportionally sampled across English, Spanish, and Italian) was independently re-coded by a second verifier. The two coders reached 85 % agreement at the topic-label level, with disagreements resolved through discussion. Although full manual double-coding was not feasible due to scale, this subset validation confirmed the consistent application of thematic labels across languages. Table 3 details the operationalization of key variables, clarifying how the structural, affective, and thematic dimensions of retweet behavior were measured across languages.

*Annotation and categorization*

Tweets and replies were manually annotated into three categories: collective, organizational, and individual, and further mapped to seven topics: brain drain causes, government-related, education and

opportunities, personal narratives, patriotism, mockery, and economy-related. These categories were derived inductively after closely reading the corpus and aligned with migration communication research (Nguyen, 2006; Tomić & Taylor, 2018; Vega-Muñoz et al., 2021; Chen et al., 2022; Rémy, 2023; Gjerazi, 2024). Items not fitting any category were labeled “N/A” and excluded. Table 4 shows how the seven topics map onto the existing literature and themes observed in our dataset, ensuring transparency in topic definition and interpretation. The data collection process is documented in Appendix A1.

*Preprocessing*

Each dataset was analyzed in its original language to preserve semantic nuance, consistent with Santoveña-Casal et al.’s (2021) recommendations on language-specific lexicon accuracy (Charquero-Ballester

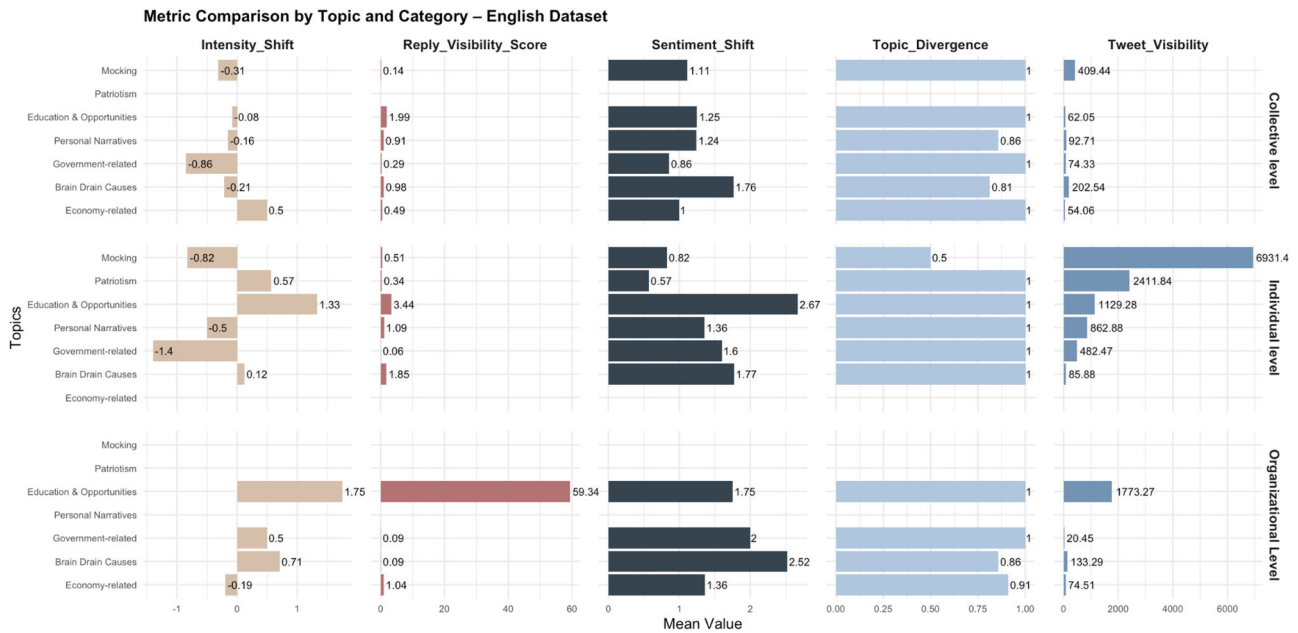


Fig. 3a. Comparative visibility, sentiment shift, and divergence patterns in English.

Note. Visibility and sentiment indices (Tweet Visibility, Reply Visibility Score, Sentiment Shift, and Intensity Shift) are standardized within language. Topic Divergence is a binary indicator and not standardized.

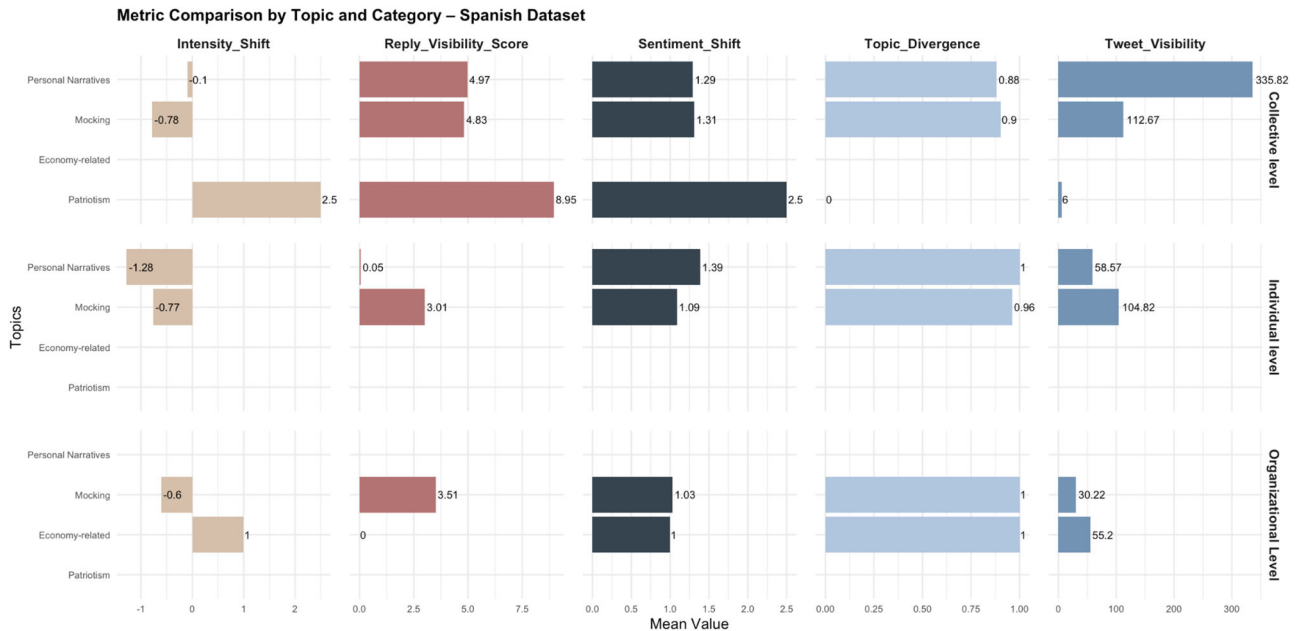


Fig. 3b. Comparative visibility, sentiment shift, and divergence patterns in Spanish.

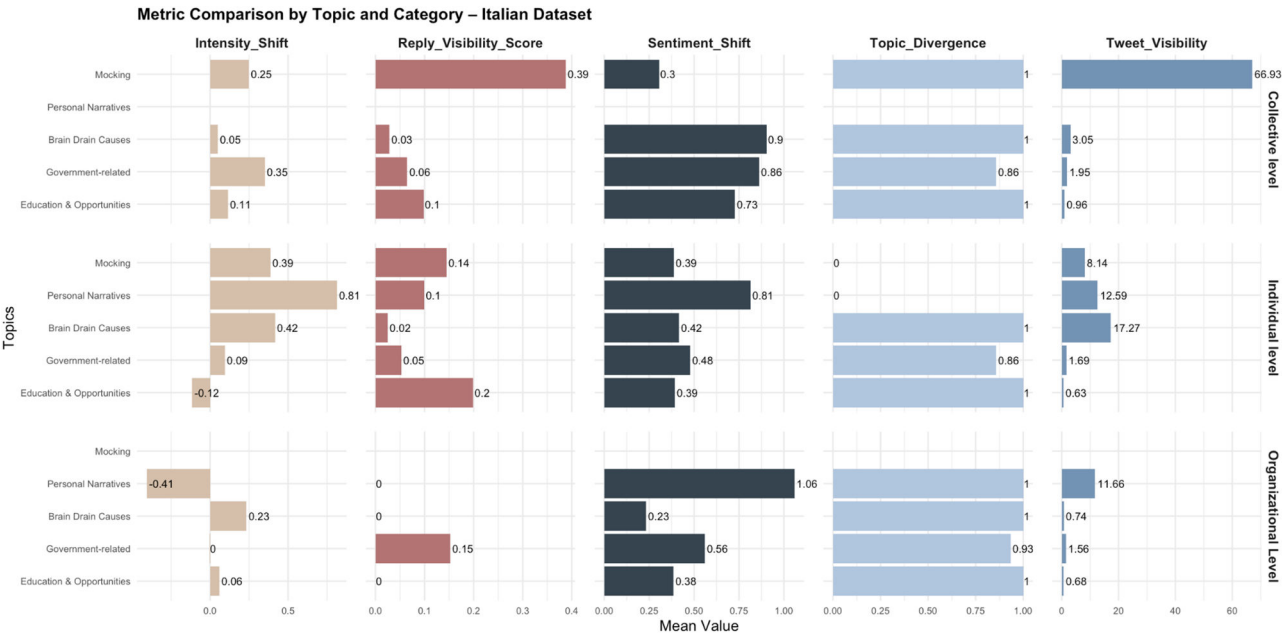
Note. Visibility and sentiment indices (Tweet Visibility, Reply Visibility Score, Sentiment Shift, and Intensity Shift) are standardized within language. Topic Divergence is a binary indicator and not standardized.

et al., 2024). Preprocessing involved lowercasing, tokenization, and stopword removal using language-specific resources: *SnowballC* for English and Spanish (Bouchet-Valat, 2020) and a curated Italian stopword list from the *Genediazjr stopwords* GitHub repository (Genediazjr, 2017). While the *Snowball* stopword lists in R were sufficiently precise for English and Spanish (Table 5), the Italian list proved less reliable. To address this, we employed the curated GitHub list, ensuring culturally appropriate filtering of high-frequency function words while retaining semantically relevant terms (Carlini et al., 2020; Ruii & Regnedda, 2022). Additional resources were integrated for sentiment analysis, including the AFINN lexicon (Nielsen, 2011), Bing lexicon (Hu & Liu,

2004), Sentix for Italian (Polignano et al., 2019), and the Emoji Sentiment Ranking dataset (Novak et al., 2015). Platform artifacts (t.co, https, amp) were removed; care was taken to avoid eliminating semantically relevant migration terms (Duong & Nguyen-Thi, 2021; Karamouz et al., 2022).

#### Sentiment analysis

Sentiment polarity was computed using validated lexicons for each language: the AFINN lexicon for English, a community-adapted AFINN (lexico\_afinn.csv) for Spanish, and Sentix for Italian (rescaled to the



**Fig. 3c. Comparative visibility, sentiment shift, and divergence patterns in Italian.**  
Note. Visibility and sentiment indices (Tweet Visibility, Reply Visibility Score, Sentiment Shift, and Intensity Shift) are standardized within language. Topic Divergence is a binary indicator and not standardized.

**Table 3**  
Operationalization of structural, affective, and thematic variables.

Variable	Definition	Operationalization
Total Retweet Impact	Combined diffusion effect of a tweet and its replies	Sum of retweets of the original tweet and retweets of its replies
Tweet Visibility	Predictive score of a tweet's structural reach	Weighted combination of tweet likes and replies, with weights estimated via PLS regression per language
Reply Visibility	Predictive score of a reply's structural reach	Weighted combination of reply likes and second-order replies, with weights estimated via PLS regression per language
Sentiment Shift	Directional change in sentiment between tweet and reply	Reply sentiment – tweet sentiment (English: Bing/AFINN; Spanish: AFINN; Italian: Sentix; emoji sentiment averaged with text polarity)
Absolute Sentiment Shift (Emotional Swing)	Magnitude of sentiment change, ignoring direction	Reply sentiment – tweet sentiment
Intensity Shift (Strength Swing)	Change in emotional magnitude regardless of polarity	Reply sentiment  –  tweet sentiment
Topic Divergence	Binary indicator of topic-level drift	1 if reply's dominant LDA topic ≠ tweet's topic; 0 otherwise
Thematic Divergence (Thematic Swing)	Continuous measure of semantic drift	1 – cosine similarity between TF-IDF vectors of tweet and reply

Note. Example when Tweet sentiment = +5 and Reply sentiment = -5 →. Sentiment shift = -10 (directional change toward negativity). Absolute sentiment shift = 10 (magnitude of emotional swing). Intensity shift = 0 (emotional strength unchanged, equally strong but opposite polarity).

AFINN range of -5 to +5). Emoji polarity scores (Novak et al., 2015) were integrated to capture multimodal expression, and the sentiment score of each message was calculated as the mean of text- and emoji-based polarity. To adapt sentiment scoring to the migration discourse, the Spanish AFINN lexicon was extended to include migration-specific terms such as *migración*, *exilio*, *fuga*, *retorno*, *talento*, *expulsión*, *refugiado*, *diáspora*, *frontera*, *xenofobia*, *asilo*, *destierro*, *barreras*, *nación*, and *país*. These words frequently appear in migration narratives and often carry negative polarity associated with loss, displacement, or constraint. Validation against the baseline lexicon produced a small negative correlation ( $r = -0.07$ , 95 % CI [-0.10, -0.04],  $n = 4\,507$ ). This modest inverse association reflects the introduction of systematically negative migration terms—such as *migración*, *fuga*, *expulsión*, and *refugiado*—which moved previously neutral or mildly positive messages toward more negative valence (Appendix Fig. A1). This divergence does not signal noise or inconsistency; rather, it shows that the adaptation captures a real, context-bound emotional tone that the generic lexicon misses—the anxiety, loss, or structural critique embedded in “fuga de cerebros,” for example. The result indicates that the extended lexicon captures new, domain-relevant emotional cues rather than random noise, improving sensitivity to the specific affective language surrounding the brain-drain debate.

From these scores, three relational measures were derived. Sentiment shift was defined as the reply's sentiment minus the tweet's sentiment (Reply – Tweet), capturing the directional change in tone. Absolute sentiment shift was defined as the magnitude of this change; it ignores direction and measures how far removed the reply is from the emotion of the original post. Intensity shift measured changes in emotional strength by comparing absolute sentiment scores for tweets and replies, indicating whether replies became more or less emotionally intense, regardless of polarity.

For example, if a tweet scored +5 and its reply scored -5, the sentiment shift would be -10 (a negative swing), the absolute sentiment



**Table 4**

Topic characteristics mapped to prior literature and empirical findings.

Topics	Scholars Findings on Recurring Themes	Reference	Our Findings
Brain Drain Causes	Economic instability Limited job opportunities Political turmoil Political insecurity	Tomić, C. H. & Taylor, K., 2018; Ajoseh, S. et al., 2024	Loss of skilled talented professionals Concern over national loss Intellectual migration
Government-related	Inadequate government policies Lack of support for research development Bureaucratic hurdles Government initiatives to attract expatriates	Vega-Muñoz, A. et al., 2021	Systemic inefficiencies Critiques towards specific government officials or private actors Complaints about corruption, lack of support, regulatory failures contributing to brain drain
Education & Opportunities	Disparities in educational standards between countries Limited opportunities for career advancement	Tomić, C. H. & Taylor, K., 2018; Ajoseh, S. et al., 2024	Mention of personal dreams for jobs Access to education and career growth opportunities Better opportunities abroad
Personal Narratives	Challenges and successes of emigrants Emotional and social implications of leaving one's homeland Feeling of loss Impact on local communities	Gjerazi, B., 2024; Rémy, F.C., 2023	Many unmet expectations in home countries Personal stories on why people left their home countries A lot of disappointment towards countries (both home and host) Emigration as a form of protest
Patriotism	Wish to return home if conditions improve Patriotism influences migration's decisions Trade-off between seeking better opportunities and the sense of loyalty to one's country	Nguyen, C. H., 2006	Nostalgia of home countries a calls for more opportunities to return and build a better version of home country Used to inspire hope or guilt
Mocking	Critics about government policies and societal conditions Coping mechanism to cope with frustrations about systematic issues	Gjerazi, B., 2024	Mocking of specific government officials Mocking between users for their ideas Sometimes derailing from serious topic with performative commentary
Economy-related	Remittance from abroad can increase home country's economy Brain drain hinder economic development and innovation in the home country	Oosterik, S. 2016; Vega-Muñoz, A. et al., 2021	Salaries and overall economic instability Limited financial opportunities, especially after Covid-19 Local economic conditions are push factors causing brain drain

Note. Columns align prior scholarship on recurring brain-drain themes with thematic patterns identified in this dataset.

**Table 5**

Stopword resources used for English, Spanish, and Italian preprocessing.

Language	Stopwords package used	Stopwords' examples
English	Snowball R Package	<i>i, me, my, myself, we, our, ours, ourselves, you, our, yours, yourself, yourselves, he, him, his, himself, she, her, hers</i>
Spanish	Snowball R Package	<i>de, la, que, el, en, y, a, los, del, se, las, por, un, para, con, no, una, su, al, lo</i>
Italian	Genediazjr stopwords list	<i>a, abbastanza, abbia, abbiamo, abbiano, abbiate, accidenti, ad, adesso, affinché, gl, agli, ahime, ahimè, ai, al, alcuna, alcuni, alcuno, all</i>

Note. Examples illustrate language-specific stopwords lists applied during text preprocessing.

shift would be 10 (a large emotional gap), and the intensity shift would be 0 (equal strength, opposite polarity). This distinction allowed us to capture the direction (sentiment shift), distance (absolute sentiment shift), and strength (intensity shift) of emotional modulation. A 10 % manually coded subsample confirmed strong agreement with the automated measures ( $\geq 0.78$ ).

#### Topic modeling and divergence

Latent Dirichlet Allocation (LDA) was applied separately to each language's tweet-reply corpus, using Gibbs sampling with a fixed random seed (1234) for reproducibility. The number of topics was set to seven, based on an inductive reading of the corpus and alignment with prior scholarship (Table 4). LDA produces a probability distribution of the topics for each text; for interpretability, each tweet or reply was assigned to its dominant topic (highest probability). Topic divergence was then coded as 1 when the dominant topic of a reply differed from that of its parent tweet, and 0 otherwise.

To complement this categorical measure, we also computed thematic divergence, a continuous measure of semantic drift. Tweets and replies were vectorized with TF-IDF, and divergence was defined as 1 – cosine

similarity between the tweet and reply. This dual strategy ensured that divergence was captured as discrete topical shifts and more subtle semantic departures.

To capture the cross-linguistic variation in engagement, we operationalized three swing measures derived from tweet-reply pairs: emotional swing (absolute sentiment shift,  $|\text{Reply} - \text{Tweet}|$ ), thematic swing (thematic divergence,  $1 - \text{cosine similarity of TF-IDF vectors}$ ), and strength swing (intensity shift,  $|\text{Reply Sentiment}| - |\text{Tweet Sentiment}|$ ), which were later compared across languages and visualized in Figs. 3a–c.

#### Visibility metrics

Following diffusion models (Brady et al., 2021), visibility was operationalized as a weighted combination of likes and replies. For each language, tweet visibility was estimated by partial least squares regression coefficients that predicted retweet counts from likes and replies. Reply visibility was computed analogously, based on likes and replies to replies. A total retweet impact measure summed retweets of original tweets and their replies. This calibration ensured that visibility reflected actual engagement pathways within each language community.

$$\text{Tweet Visibility}_i = \beta_1 \times \text{Likes}_i + \beta_2 \times \text{Replies}_i$$

$$\text{Reply Visibility}_i = \beta'_1 \times \text{Likes}_{\{r,i\}} + \beta'_2 \times \text{Replies}_{\{r,i\}} \quad (1)$$

where  $\beta_1$ ,  $\beta_2$ ,  $\beta'_1$ , and  $\beta'_2$  are language-specific regression weights estimated via ordinary least squares (OLS) (see Appendix Table A1). Both indices were standardized within each language to ensure comparability across linguistic communities.

This design avoids spurious cross-linguistic contrasts that may arise from different engagement baselines. Appendix Table A2 reports the estimated weights for each language, indicating how engagement components contribute to the Latent Visibility Index in a language-specific manner. Standardized scores were used for all subsequent

regression and interaction models, and figures were labeled “standardized within language” to avoid misinterpretation.

### Statistical analyses

The descriptive evidence in Table 2 and Figs. 1–3 already suggests systematic cross-linguistic contrasts. English activity was concentrated on collective topics, generating the highest likes per tweet and reflecting collective coherence. Spanish activity was more dispersed and reflected plurality, with organizational tweets being the most frequent, but individual topics attracting stronger engagement. Italian activity centered on organizational themes, with modest visibility, reflecting institutional optimism. Figs. 1a–b confirmed these tendencies: English clustered around *causes* and *patriotism*, Spanish spanned personal, economic, and institutional themes, and Italian emphasized *government* and *education*. Fig. 2 highlighted distinct sentiment distributions: English clustered neutral-to-positive, Spanish replies spread more widely and often negative, and Italian consistently positive. Finally, Fig. 3a–c confirmed differences in engagement: English publics displayed strong tweet visibility but low reply visibility and divergence, Spanish publics displayed strong reply visibility and higher divergence yet coherent threads, and

Italian publics showed modest visibility with consistently positive sentiment shifts.

To test whether these descriptive contrasts were statistically robust, we compared three continuous measures across languages: absolute sentiment shift (emotional swing), thematic divergence (thematic swing), and intensity shift (strength swing), using Levene’s tests, one-way ANOVA, and Tukey HSD post-hoc comparisons ( $\alpha = 0.05$ ). Levene’s tests were highly significant ( $p < .001$ ) for all three measures, reflecting unequal variances across the languages.

For absolute sentiment shift (emotional swing), ANOVA showed strong differences across languages,  $F(2, 15,754) = 4103, p < .001$ . Tukey comparisons revealed that English replies produced significantly larger swings than both Italian ( $p < .001$ ) and Spanish ( $p < .001$ ), while Spanish–Italian differences were negligible ( $p = .76$ ). This confirms that English publics engage through sharper emotional swings, often amplifying or reversing the polarity of the original tweet.

For thematic divergence (thematic swing), ANOVA also indicated significant differences,  $F(2, 15,754) = 64.79, p < .001$ . Post-hoc comparisons showed that Spanish replies were significantly more convergent (lower divergence) than both English ( $p < .001$ ) and Italian ( $p < .001$ ), while English–Italian differences were not significant ( $p = .36$ ).

**Table 6**  
Regression models predicting total retweet impact (thematic divergence specification).

Panel A: English								
Predictors	Model 3		Model 4		Model 5		Model 6	
	Estimates	p	Estimates	p	Estimates	p	Estimates	p
(Intercept)	−313.59	0.029	−283.61	0.047	−278.11	0.052	901.21	<0.001
Tweet Visibility	1.03	<0.001	1.03	<0.001	1.01	<0.001	−0.67	<0.001
Reply Visibility	0.88	<0.001	0.87	<0.001	0.83	<0.001	59.86	0.007
Sentiment Shift	−5.94	0.814	−6.76	0.788	−99.69	0.006		
Thematic Divergence	286.42	0.06	103.27	0.503	257.54	0.09	−1134.2	<0.001
Categories replies- [Individual level]			171.51	<0.001			157.2	<0.001
Categories replies- [Collective level]			149.1	<0.001			137.94	<0.001
Tweet Visibility × Sentiment Shift					0.12	<0.001		
Reply Visibility × Thematic Divergence							−62.8	0.008
Tweet Visibility × Thematic Divergence							1.79	<0.001
Observations	3073		3073		3073		3073	
R <sup>2</sup> / R <sup>2</sup> adjusted	0.817 / 0.817		0.819 / 0.819		0.818 / 0.817		0.825 / 0.824	
Note. All continuous predictors (Visibility, Sentiment Shift, and Thematic Divergence) are standardized within language. Estimates are unstandardized coefficients.								
Panel B: Spanish								
Predictors	Model 3		Model 4		Model 5		Model 6	
	Estimates	p	Estimates	p	Estimates	p	Estimates	p
(Intercept)	532.89	<0.001	643.15	<0.001	472.27	0.001	333.76	0.111
Tweet Visibility	0.79	<0.001	0.79	<0.001	0.84	<0.001	1.43	<0.001
Reply Visibility	0.99	<0.001	0.99	<0.001	0.98	<0.001	−7.54	<0.001
Sentiment Shift	106.8	0.421	104.59	0.43	−864.19	<0.001		
Thematic Divergence	−322.06	0.041	−261.11	0.1	−278.45	0.075	72.15	0.744
Categories Individual level			−203.19	0.002			−190.84	0.003
Categories Collective level			−169.34	0.003			−167.74	0.003
Tweet Visibility × Sentiment Shift					1.75	<0.001		
Reply Visibility × Thematic Divergence							8.58	<0.001
Tweet Visibility × Thematic Divergence							−0.69	0.02
Observations	4490		4490		4490		4490	
R <sup>2</sup> / R <sup>2</sup> adjusted	0.570 / 0.570		0.571 / 0.571		0.576 / 0.576		0.584 / 0.583	
Note. All continuous predictors (Visibility, Sentiment Shift, and Thematic Divergence) are standardized within language. Estimates are unstandardized coefficients.								
Panel C: Italian								
Predictors	Model 3		Model 4		Model 5		Model 6	
	Estimates	p	Estimates	p	Estimates	p	Estimates	p
(Intercept)	−2.61	0.706	−2.44	0.72	−3.02	0.664	−5.52	0.572
Tweet Visibility	0.99	<0.001	0.99	<0.001	0.99	<0.001	1.06	<0.001
Reply Visibility	1.31	0.603	0.4	0.873	1.4	0.58	20.95	0.579
Sentiment Shift	17.14	<0.001	16.76	<0.001	17.99	<0.001		
Thematic Divergence	4.4	0.545	7.44	0.303	4.86	0.506	11.21	0.281
Categories Individual level			−4.23	0.002			−4.34	0.002
Categories Collective level			−3.14	0.013			−3.91	0.003
Tweet Visibility × Sentiment Shift					−0.05	0.519		
Reply Visibility × Thematic Divergence							−21.77	0.59
Tweet Visibility × Thematic Divergence							−0.08	0.776
Observations	253		253		253		253	
R <sup>2</sup> / R <sup>2</sup> adjusted	0.981 / 0.981		0.982 / 0.982		0.981 / 0.981		0.981 / 0.980	

Note. All continuous predictors (Visibility, Sentiment Shift, and Thematic Divergence) are standardized within language. Estimates are unstandardized coefficients.

This suggests that Spanish publics, while branching into diverse topical categories, sustain low thematic swings by maintaining coherence after threads are established.

For intensity shift (strength swing), the ANOVA showed moderate but significant differences,  $F(2, 15,754) = 42.55, p < .001$ . English replies showed greater changes in emotional strength compared to Italian ( $p = .024$ ) and Spanish ( $p < .001$ ), while Spanish–Italian differences were minimal ( $p = .75$ ). This indicates that English publics are more likely to alter the strength of emotional expression, either intensifying or dampening the tone, whereas Spanish and Italian publics maintain more stable emotional magnitudes.

Together, these results show that engagement dynamics vary by language along three distinct axes of swing. English publics diverge most in emotional and strength swings, reflecting engagement through the emotional modulation of coherent collective narratives. Spanish publics minimize both emotional and strength swings, while also showing low thematic swings, reflecting plurality across topics but coherence within threads. Italian publics sustain optimism with low emotional and strength swings, but moderate thematic swings, reflecting replies that remain consistently positive even when the topics shift slightly.

These statistical contrasts validate the descriptive insights from Table 2 and Figs. 1–3, and establish the empirical foundation for the regression analyses (Tables 6 and 7), which examine how visibility, sentiment, and divergence jointly shape engagement outcomes across languages.

## Results

### Descriptive patterns and swing analysis

The baseline descriptives (Table 2; Figs. 1–3) reveal distinct cultural orientations in the brain drain discourse. English activity clustered in collective topics such as the causes of brain drain and patriotism, producing the highest likes per tweet and emphasizing coherent societal narratives. Spanish activity reflected plurality and was broadest in organizational tweets, but gained the most traction from individual stories. Italian discourse centered on organizational themes such as government and education, but overall engagement was modest, consistent with institutional optimism. Sentiment distributions (Fig. 2) reinforced these contrasts: English replies were clustered around a neutral-to-positive tone, Spanish replies were more dispersed and often negative, and Italian replies were consistently positive. Fig. 3 confirms these dynamics structurally, showing that English engagement hinged on tweet visibility, Spanish on reply visibility and divergence, and Italian on positive sentiment shifts.

Continuous “swing” measures further validated these descriptive patterns. Emotional swings, measured as absolute sentiment shift, were the largest in English, underscoring affective volatility. Thematic swings, measured as cosine-based divergence, were lowest in Spanish, suggesting greater topical alignment despite plurality. Strength swings, measured as intensity shifts in emotional magnitude, were modest overall but slightly greater in English than Spanish or Italian. Taken together, these descriptive and continuous indicators highlight three distinct logics: English publics prioritize coherence, Spanish publics sustain plurality, and Italian publics prefer to engage in optimism.

### Regression results: thematic divergence as the main specification

Table 6 shows the main regression models predicting total retweet impact, where divergence is operationalized continuously as the thematic distance between tweet and reply. Across all languages, tweet visibility was the strongest positive predictor of diffusion, though the role of other factors diverged by language. In English (Panel A), engagement was driven primarily by tweet visibility ( $\beta \approx 1.0, p < .001$ ), with reply visibility also contributing positively. The interaction models revealed that divergence was penalized when replies were highly visible

**Table 7**

Regression models predicting total retweet impact (topic divergence specification).

Panel A: English				
	Model 1	Model 2	Model 3	Model 4
(Intercept)	40.83 (117.68)	46.31 (130.29)	18.61 (101.62)	−46.41 (132.07)
Tweet Visibility	1.00 *** (0.02)	1.00 *** (0.02)	0.80 *** (0.04)	0.63 *** (0.02)
Reply Visibility	0.78 (2.49)	0.47 (2.57)	7.83 *** (2.55)	9.22 (43.06)
Sentiment Shift	6.58 (6.08)	6.04 (17.78)	−28.80 (16.63)	96.61 (67.25)
Topic Divergence	−50.39 * (118.93)	−22.52 (122.52)	−47.68 (104.39)	133.77 (133.77)
Categories Individual level		−95.72 ** (78.34)		88.52 (15.07)
Categories Collective level		−14.21 (78.23)		13.07 (67.25)
Tweet Visibility × Sentiment Shift			0.14 *** (0.03)	
Reply Visibility × Topic Divergence				0.21 * (43.26)
Tweet Visibility × Topic Divergence				−0.35 ** (0.06)
R <sup>2</sup>	0.97	0.97	0.98	0.98
Adj. R <sup>2</sup>	0.97	0.97	0.98	0.98
Num. obs.	78	78	78	78
Note. All variables are standardized within language. Topic Divergence is a binary indicator derived from dominant Latent Dirichlet Allocation (LDA) topics. Estimates are unstandardized coefficients. *** $p < 0.001$ ; ** $p < 0.01$ ; * $p < 0.05$				
Panel B: Spanish				
	Model 1	Model 2	Model 3	Model 4
(Intercept)	56.72 (66.48)	35.94 (110.38)	56.18 (66.52)	7.02 (115.26)
Tweet Visibility	0.98 *** (0.07)	0.98 *** (0.07)	0.99 *** (0.07)	1.84* (0.82)
Reply Visibility	−1.30 (0.95)	−1.29 (0.96)	−1.36 (0.96)	−8.43 (6.03)
Sentiment Shift	−0.60 (11.57)	−0.88 (11.69)	−5.29 (12.29)	(36.55)
Topic Divergence	23.91 (68.68)	28.48 (69.66)	25.20 (68.65)	88.42 (88.89)
Categories Individual level		6.47 (91.44)	(25.20)	88.42 (23.48)
Categories Collective level		23.49 (88.78)		(88.78)
Tweet Visibility × Sentiment Shift				66.50 (66.50)
Reply Visibility × Topic Divergence			0.04 (0.04)	
Tweet Visibility × Topic Divergence				13.65* (6.50)
				−0.85 (0.82)
R <sup>2</sup>	0.60	0.60	0.60	0.62
Adj. R <sup>2</sup>	0.59	0.58	0.59	0.62
Num. obs.	181	181	181	181
Note. All variables are standardized within language. Topic Divergence is a binary indicator derived from dominant Latent Dirichlet Allocation (LDA) topics. Estimates are unstandardized coefficients. *** $p < 0.001$ ; ** $p < 0.01$ ; * $p < 0.05$				
Panel C: Italian				
	Model 1	Model 2	Model 3	Model 4
(Intercept)	0.68 (1.28)	0.94 (1.75)	0.94 (1.27)	−2.69 (2.53)
Tweet Visibility	1.00 *** (0.02)	1.00 *** (0.02)	1.21 *** (0.08)	2.30 *** (0.36)
Reply Visibility	−0.56 (2.49)	−0.64 (2.54)	−0.89 (2.44)	−1.46 (23.85)

(continued on next page)

Table 7 (continued)

Panel A: English				
	Model 1	Model 2	Model 3	Model 4
Sentiment Shift	-0.18 (0.50)	-0.16 (0.51)	0.69 (0.59)	-0.43
Topic Divergence	1.35 * (0.44)	0.41 (0.44)	0.24 (0.24)	(0.59)
Categories Individual level		-0.13 (1.39)		-2.69 (0.59)
Categories Collective level		-0.43 (1.31)		0.47 (0.24)
Tweet Visibility × Sentiment Shift			-0.37 * (0.14)	
Reply Visibility × Topic Divergence				0.76 (23.95)
Tweet Visibility × Topic Divergence				-1.30 (0.36)
R <sup>2</sup>	0.96	0.96	0.96	0.97
Adj. R <sup>2</sup>	0.96	0.96	0.96	0.67
Num. obs.	121	121	121	121

Note. All variables are standardized within language. Topic Divergence is a binary indicator derived from dominant Latent Dirichlet Allocation (LDA) topics. Estimates are unstandardized coefficients.

\*\*\*  $p < 0.001$ .

\*\*  $p < 0.01$ .

\*  $p < 0.05$ .

(Reply Visibility × Thematic Divergence = -62.8,  $p = .008$ ), while sentiment shifts were tolerable only when buffered by high tweet visibility (Tweet Visibility × Sentiment Shift = 0.12,  $p < .001$ ). In Spanish (Panel B), both tweet and reply visibility predicted diffusion, but with different dynamics. Divergent replies reduced engagement (-322,  $p = .041$ ), yet this penalty was significantly cushioned when reply visibility was high (Reply Visibility × Thematic Divergence = +8.58,  $p < .001$ ). Simultaneously, large negative sentiment shifts destabilized engagement (-864,  $p < .001$ ), suggesting that plurality is structurally tolerated

but emotionally fragile. In Italian (Panel C), sentiment shift was the dominant predictor ( $\beta \approx 17$ ,  $p < .001$ ), visibility played a weaker role, and thematic divergence was non-significant, underscoring that affective tone rather than structural coherence sustains engagement.

All models were tested for multicollinearity and heteroskedasticity. Variance Inflation Factors (VIFs) across all specifications remained below 3.0 (maximum = 2.75), indicating no multicollinearity concerns. To ensure robustness, each model was re-estimated using heteroskedasticity-consistent (HC3) standard errors. The coefficients and significance levels remained stable, confirming that the main results were not sensitive to heteroskedasticity.

The visibility coefficients remained large and significant across languages. In the main specification (Table 6), a one-standard-deviation increase in tweet visibility predicted roughly a one-unit increase in total retweet impact for English and Spanish, and a slightly stronger 1.1-unit increase for Italian. Reply visibility also showed consistent positive effects, suggesting that conversational amplification—when replies attract attention—reinforces overall diffusion.

The robustness specification (Appendix Table A3) yielded nearly identical estimates, with minor increases in standard errors under HC3 correction. These results indicate that the visibility indices are statistically stable and substantively interpretable: greater visibility reliably translates into higher engagement, even when accounting for sentiment and thematic divergence.

#### Interaction effects

Figs. 4a–c present the interaction effects, showing how visibility fundamentally conditions the interpretation of sentiment and thematic divergence across languages.

In English (Fig. 4a), sentiment shifts suppressed engagement under low visibility, yet were buffered once the tweets gained prominence, suggesting that publics tolerate emotional fluctuation insofar as it remains anchored in structurally visible posts. Visibility therefore functions as a stabilizing filter: emotional volatility is tolerated when expressed through highly visible messages yet discouraged in peripheral

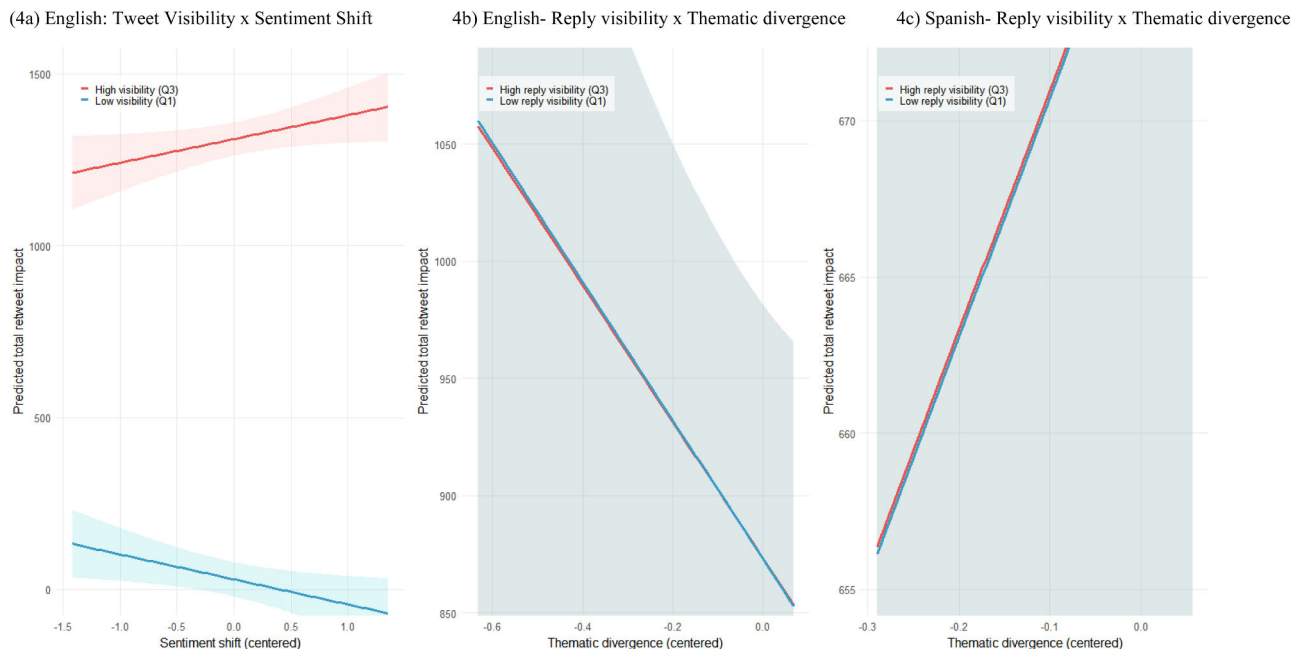


Fig. 4. Interaction effects of visibility, sentiment shift, and divergence across languages.

(4a) English: Tweet Visibility x Sentiment Shift (4b) English- Reply visibility x Thematic divergence (4c) Spanish- Reply visibility x Thematic divergence

Note. Panels (4a–4c) depict regression-derived interaction effects between visibility and sentiment/thematic divergence for English and Spanish. All variables are standardized within language. Lines represent predicted retweet impact for high (Q3) and low (Q1) visibility. High reply visibility boosts diffusion for coherent English threads but loses advantage when replies diverge, whereas in Spanish discussions, visibility penalties diminish as conversations diversify.



or low-visibility contexts. Similarly, as shown in Fig. 4b, highly visible replies were penalized when they drifted thematically ( $\beta = -62.8, p = .008$ ), indicating that visibility amplifies the reward for coherence while dampening the diffusion of off-topic engagement. In the English context, visibility thus consolidates emotional and thematic order—reinforcing coherence as a normative ideal for public discourse.

In Spanish (Fig. 4c), reply visibility initially reduced engagement ( $\beta = -7.54, p < .001$ ), yet this penalty was progressively mitigated as thematic divergence increased ( $\beta = +8.58, p < .001$ ). The parallel upward slopes in Fig. 4c indicate that heterogeneous discussions regained engagement even under high visibility, effectively transforming divergence from a liability into a participatory resource. In contrast to the English pattern—where visibility rewards coherence—Spanish publics appear to use visibility to normalize diversity, interpreting multiple viewpoints as a sign of authentic deliberation rather than fragmentation. However, large negative sentiment shifts continued to destabilize engagement, underscoring the fragile equilibrium between structural inclusivity and affective volatility.

Italian publics showed a distinct pattern: thematic divergence effects were statistically non-significant, whereas sentiment optimism consistently amplified engagement regardless of visibility. This suggests that Italian engagement dynamics rely more on affective positivity than on structural visibility or thematic alignment. Rather than moderating emotional volatility or reinforcing thematic coherence, Italian users appear to reward positive tone itself as the principal signal of conversational value.

Taken together, these results reveal three differentiated visibility logics. English publics use visibility to consolidate coherence, stabilizing attention around emotionally and thematically consistent narratives. Spanish publics use visibility to legitimize divergence, transforming heterogeneity into a socially recognized form of engagement. Italian publics, by contrast, favor engagement in affective positivity rather than structural visibility or thematic order.

These cross-linguistic patterns demonstrate that visibility does more than amplify exposure—it functions as a culturally adaptive amplifier, translating platform affordances into locally meaningful forms of participation. The interplay between sentiment, divergence, and visibility thus reflects not only communicative strategies but also underlying cultural preferences for coherence, plurality, or optimism, which will be discussed in the following section.

#### *Complementary robustness check: topic divergence*

To ensure that these findings are not artifacts of how divergence is defined, Table 7 re-estimates the models using Topic Divergence, a binary indicator of whether replies switch dominant topics relative to the tweet. This specification complements the continuous semantic measure in Table 6 by capturing sharper categorical breaks in the conversation. The results broadly converge with the main specification while adding nuance. In English, topic divergence reduced baseline amplification ( $\beta = -50.39, p < .05$ ) yet recovered when divergent replies themselves gained visibility, whereas high tweet visibility intensified penalties. Spanish models again showed that reply visibility legitimized divergence ( $\beta = 13.65, p < .05$ ), reproducing the contingency observed with thematic divergence, although topic switching was penalized in already high-visibility threads. The Italian results remained stable, with topic divergence showing small or non-significant effects, whereas sentiment alignment continued to dominate engagement. We estimated topic divergence (binary) via within-language LDA and correlated it with thematic divergence ( $1 - \text{cosine similarity}$ ). The correlations were positive but small in English ( $r = 0.12$ ) and Spanish ( $r = 0.05$ ), and modest in Italian ( $r = 0.31$ ), indicating that topic switching (categorical drift) and semantic distance (continuous drift) capture related yet non-identical aspects of conversational change. Results were robust to nonparametric (Spearman) estimates and reported with 95 % CIs in Table A3.

Taken together, the robustness check confirms that visibility remains the central driver across languages; sentiment shift operates differently by cultural context; and divergence matters only under conditions of visibility. Thematic divergence provides a continuous, language-agnostic measure of semantic drift, whereas topic divergence highlights sharper conversational ruptures. The use of both operationalizations strengthens the credibility of the results and clarifies why Spanish and English publics exhibit shifting penalties and buffers depending on the definition of divergence, whereas Italian publics remain consistently governed by affective framing.

## **Discussion**

This study examined how multilingual publics engage with brain drain debates on X, focusing on English, Spanish, and Italian discourse. By analyzing visibility, sentiment, and divergence, we identified not only surface-level differences in activity, but also deeper cultural logics that shape how online publics respond to migration crises. The main specification, based on thematic divergence, showed clear cross-linguistic contrasts, whereas the robustness check with topic divergence confirmed these dynamics under an alternative definition of conversational drift. Together, these models provide convergent evidence that visibility, sentiment, and divergence interact differently across linguistic communities, reinforcing that cultural logics of engagement are not artifacts of measurement choice but robustly observed features of online discourse.

The results highlight three orientations. English publics emphasize coherence: emotional variation is tolerated only when attached to highly visible tweets, whereas divergence, whether measured thematically or topically, is penalized once it becomes visible. Spanish publics value diversity of voice: divergent perspectives tend to reduce engagement when hidden, yet become legitimate and even rewarding when reply visibility is high. This result holds under both divergence measures. Simultaneously, negative emotions destabilize plurality, underscoring the fragility of affective dynamics. Italian publics rely on optimism: positive sentiment shifts consistently amplify engagement, and divergence plays a small role regardless of how it is defined. These findings underscore that bonding, bridging, and affective capital operate differently across linguistic communities, shaping how the migration discourse is interpreted and diffused.

Theoretically, this contributes to research on social capital by showing how bonding and bridging mechanisms take on language-specific forms, and by introducing affective capital as a complementary lens. In English debates, visibility consolidates coherence, reinforces bonding capital by privileging thematic alignment, and penalizes drift. In Spanish debates, reply visibility cushions divergence and aligns with bridging capital by granting legitimacy to diverse perspectives, even as negativity threatens stability. Italian debates point to affective capital, where optimism provides continuity regardless of structural cues, consistent with recent findings on the role of institutional trust in shaping Southern European migration narratives (Patel & Kim, 2025). By using both semantic (thematic) and categorical (topic) measures of divergence, the study demonstrates that these theoretical insights are not tied to a single operationalization but hold across complementary definitions.

Methodologically, this study demonstrates the value of combining multiple computational approaches for comparative migration research. The integration of sentiment scoring, swing measures, and interaction modeling allowed us to capture dynamics that are difficult to detect through manual coding alone. By using both thematic and topic divergence, we detected both subtle semantic drifts and sharper categorical breaks, showing that the interpretation of divergence itself is conditioned by visibility. This dual operationalization directly addresses concerns about definitional clarity and strengthens the reliability of cross-linguistic comparisons.



**Table 8**  
Policy implications of multilingual interaction effects.

Insights	Empirical Result	Interpretive Meaning	Policy / Communication Action
How does visibility shape engagement?	English: Tweet visibility buffers emotional swings (Fig. 4a); Spanish: Reply visibility cushions divergence (Fig. 4b); Italian: Sentiment shifts dominate over visibility.	Visibility functions differently across publics: coherence under English, plurality under Spanish, optimism under Italian.	English: Prioritize making coherent policy narratives highly visible (e.g., government accounts amplifying unified Brexit migration updates). Spanish: Ensure diverse replies are surfaced (e.g., community organizations retweeting personal migration stories). Italian: Use visibility less for reach, more to showcase positive tone (e.g., ministries amplifying stories of opportunity programs).
How do sentiment shifts affect diffusion?	English: Large emotional swings penalized unless buffered by visibility; Spanish: Negative shifts destabilize engagement; Italian: Positive sentiment shifts amplify retweets (Table 6).	Affect is culturally filtered: emotional restraint in English, volatility in Spanish, optimism in Italian.	English: Avoid overly emotional framings unless backed by authoritative accounts. Spanish: Anticipate volatility; provide empathetic framing when negative emotions spike. Italian: Lean on affective optimism (e.g., highlighting education/job initiatives).
How is thematic divergence interpreted?	English: Divergent replies penalized when visible (Fig. 4c); Spanish: Divergence reframed as legitimate under reply visibility; Italian: Divergence effects nonsignificant.	Divergence = coherence violation in English, inclusivity marker in Spanish, neutral in Italian.	English: Keep messaging thematically consistent; manage reply threads to reduce off-topic drift (e.g., official FAQs responding directly). Spanish: Use divergent voices to broaden legitimacy (e.g., retweeting varied migrant experiences). Italian: Focus on optimism regardless of divergence (e.g., framing reforms positively).



**Fig. 5.** Cross-linguistic engagement swings (emotional, thematic, strength).  
*Note.* Emotional swing is measured as the absolute sentiment shift, i.e., the magnitude of change in polarity between a tweet and its reply ( $|\text{Reply Sentiment} - \text{Tweet Sentiment}|$ ). Larger values indicate greater emotional contrast. Thematic swing is measured as thematic divergence, i.e., the degree of topical drift between tweet and reply computed as  $(1 - \text{cosine similarity})$  of their TF-IDF vectors. Larger values indicate replies departing further from the original tweet's content. Strength swing is measured as intensity shift, i.e., the change in absolute sentiment magnitude ( $|\text{Reply Sentiment}| - |\text{Tweet Sentiment}|$ ). Positive values reflect replies becoming more emotionally intense than the tweet; negative values indicate dampening. All measures are standardized within language.

Finally, the findings carry direct implications for migration communication and policy. English-speaking publics appear most responsive to coherent messaging that leverages structural visibility to stabilize discourse. Spanish-speaking publics thrive on plurality, suggesting that communicators should design strategies that encourage diverse voices, while preparing for volatility when negative sentiment emerges. Italian-speaking publics respond most strongly to optimism, highlighting the importance of positive framing in sustaining institutional trust. These differentiated logics show that digital communication on migration cannot adopt a one-size-fits-all approach. Instead, strategies must adapt to the cultural frames embedded in each language, leveraging visibility, sentiment, and divergence differently to build credibility and engagement in multilingual contexts. Table 8 summarizes the study's practical implications by linking research insights to the empirical results, interpretive meaning, and policy and communication actions. Importantly, the synthesis draws on evidence from both the main specification (Table 6, Thematic Divergence) and the

complementary robustness check (Table 7, Topic Divergence), ensuring that the recommended strategies reflect findings that are robust across alternative operationalizations of divergence.

Fig. 5 presents a conceptual map of cross-linguistic engagement dynamics across three axes of conversational swing: emotional swing (absolute sentiment shift, measuring polarity contrasts between tweets and replies), thematic swing (thematic divergence, capturing topical drift between tweets and replies), and strength swing (intensity shift, reflecting changes in the magnitude of sentiment regardless of polarity). English publics exhibit high emotional and strength swings, reflecting sharp polarity contrasts and intensity modulation, but moderate thematic swings, consistent with collective coherence reinforced through emotional modulation. Spanish publics show low emotional and strength swings and the lowest thematic swing, reflecting expressive plurality across topics but coherence within reply threads. Italian publics show low emotional and strength swings but moderate thematic swings, reflecting consistent optimism in tone even when the topical

focus shifts. Together, these axes illustrate how different cultural logics shape the discourse on brain drain on X: English through the emotional modulation of collective narratives, Spanish through plural yet coherent branching, and Italian through optimistic stability.

### Conclusion, limitations, and future research

This study shows that multilingual publics interpret and engage with the debate on migration through distinct cultural logics. English publics favor coherence, Spanish publics sustain plurality, and Italian publics rely on optimism. The amenability of these patterns to the use of computational methods allowed us to trace how visibility, sentiment, and divergence interact across thousands of tweets and replies. By integrating sentiment analysis, thematic divergence, and interaction modeling, this study demonstrates how computational approaches can uncover the dynamics of coherence, plurality, and optimism that are otherwise difficult to observe.

Simultaneously, the study's limitations must be acknowledged. The reliance on X's API may have introduced a sampling bias, as the platform's public data stream is not fully representative of all users or tweets, and patterns of activity differ substantially across linguistic communities. This unevenness may not only affect the relative weight of English, Spanish, and Italian debates in our dataset but also the dynamics of visibility captured in our models. Moreover, while our analysis focused on visibility, sentiment, and divergence, other powerful drivers of migration debates, such as explicit references to violence or insecurity, were not modeled directly, even though they remain central to the lived experience of mobility. Sentiment scoring relied on language-specific lexicons, which were rescaled and validated but are ultimately unable to fully capture the cultural nuance of irony, sarcasm, or context-dependent meanings. Thematic divergence was measured using TF-IDF cosine similarity and topic modeling, both robust methods but limited compared to newer embedding-based models that can better capture semantic subtleties. In addition, our analysis was conducted at the comment level rather than the user level, which prevented us from connecting findings to network measures of centrality,<sup>2</sup> bridging, or influence; the same user rarely appears across threads, limiting actor-level inferences. While this study focused on comment-reply dynamics rather than user-level centrality or modularity, future research could extend our findings by mapping influential users, their bridging roles, and structural positions in multilingual debates. Such analyses would complement our focus on discourse-level mechanisms by linking narrative dynamics to the architecture of online networks. Finally, the design remains observational, identifying associations rather than causal effects. In summary, technology can only get us so far: computational tools are powerful for revealing structural and affective dynamics in digital publics, but they cannot fully resolve questions of context, causality, or interpretation without complementary approaches.

Future research can address these limitations by expanding the linguistic and platform scope of analysis, incorporating multimodal

environments such as TikTok or YouTube, and extending beyond European languages to capture a wider range of cultural frames. Advances in transformer-based sentiment models and multilingual embeddings could further improve semantic accuracy, while longitudinal and experimental designs would help clarify causal mechanisms; for example, testing whether visibility interventions or framing strategies stabilize or destabilize discourse. Integrating network-level analysis would also allow scholars to connect content patterns to actor influence, bridging dynamics, and eigenvector centrality. Importantly, linking online discourse more directly to offline drivers of migration, such as violence or economic insecurity, would further ground computational analysis in lived realities.

Analyses were conducted on publicly accessible posts from X using the Academic Research API, in full compliance with the platform's developer policy. All analyses were performed at the sentence-reply level; no user identifiers, handles, or tweet IDs were stored or analyzed. The data were de-identified, and all scripts operated on synthetic examples to demonstrate workflow reproducibility. A minimal replication package (Park, 2025)—including preprocessing, sentiment, divergence, modeling code, and synthetic tweet-reply examples—is hosted on the OSF at <https://doi.org/10.17605/OSF.IO/S6VZ2>. This archive enables full pipeline reproduction without distributing original user content and maintains both ethical and legal compliance with data-sharing norms.

Despite these limitations, this study demonstrates the value of computational social science for migration research. By combining theory-driven categories with computational measures, we show how digital publics negotiate coherence, plurality, and optimism across languages.

The implications for policymakers and communicators are: multilingual communication strategies must be adaptive and responsive to the cultural logics embedded in language, instead of assuming uniform digital publics. Table 8 summarizes these implications by connecting the research questions to the empirical results, theoretical meaning, and policy recommendations. Simultaneously, our results highlight both the promise and limits of computational analysis. Technology can only get us so far: while large-scale data and automated measures illuminate structural and affective dynamics that would otherwise remain hidden, a deeper understanding of migration discourse requires the combination of these tools with contextual, qualitative, and causal approaches.

### CRedit authorship contribution statement

**Jimi Park:** Writing – review & editing, Writing – original draft, Validation, Supervision, Software, Resources, Methodology, Investigation, Formal analysis, Conceptualization. **Nicole Pitassi:** Writing – review & editing, Writing – original draft, Visualization, Data curation.

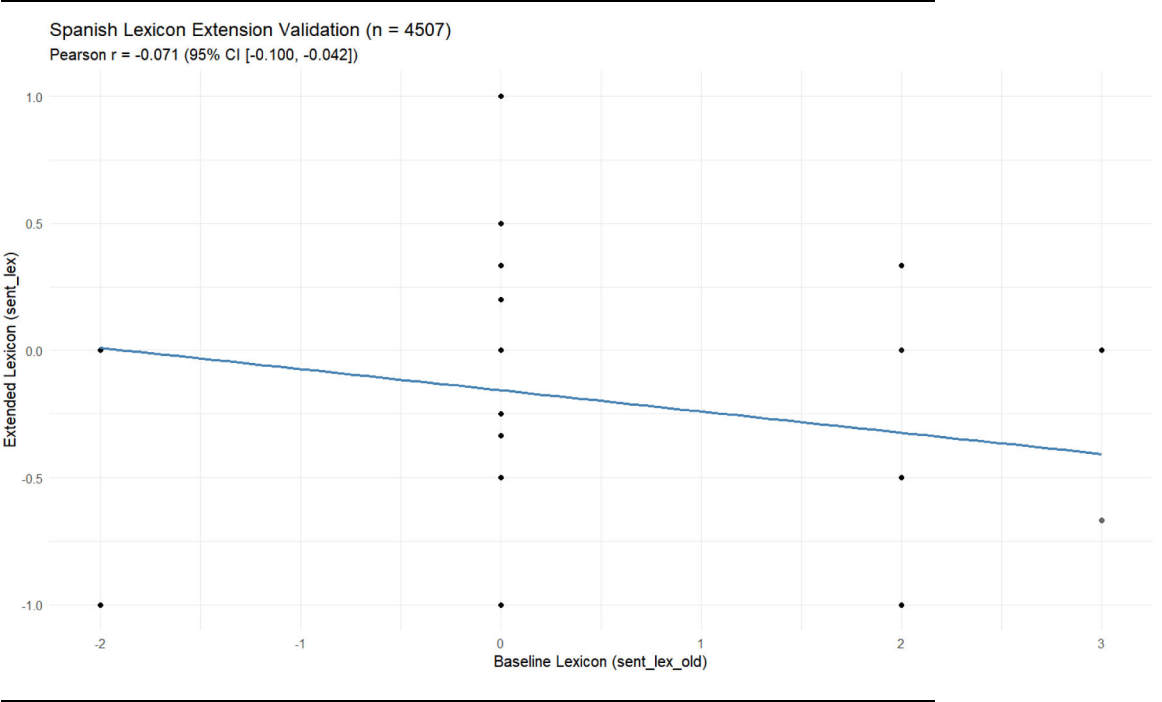
### Declaration of competing interest

None.

<sup>2</sup> We thank the anonymous reviewer for suggesting the incorporation of quantitative network metrics.

Appendix A1

**Table A1**  
Data collection workflow and topic justification.



*Note.* The number of topics was set to seven, based on both inductive reading of the corpus and alignment with prior scholarship (Table 4).

**Table A2**  
Cross-linguistic visibility weights and mechanisms.

Language	Tweet $\beta_1$ (Replies)	Tweet $\beta_2$ (Likes)	Reply $\beta'_1$ (Likes)	Reply $\beta'_2$ (Replies)	Dominant Mechanism	Social-Capital Dimension	Interpretation
English	-0.04	0.27	0.12	0.06	Endorsement-driven diffusion	Bonding	Likes amplify collective coherence and positive alignment.
Spanish	1.99	0.08	0.15	1.02	Dialogue-based diffusion	Bridging	Replies and re-replies sustain discursive plurality.
Italian	0.60	0.11	0.05	0.00	Balanced diffusion	Affective	Emotional resonance and reciprocity drive visibility.

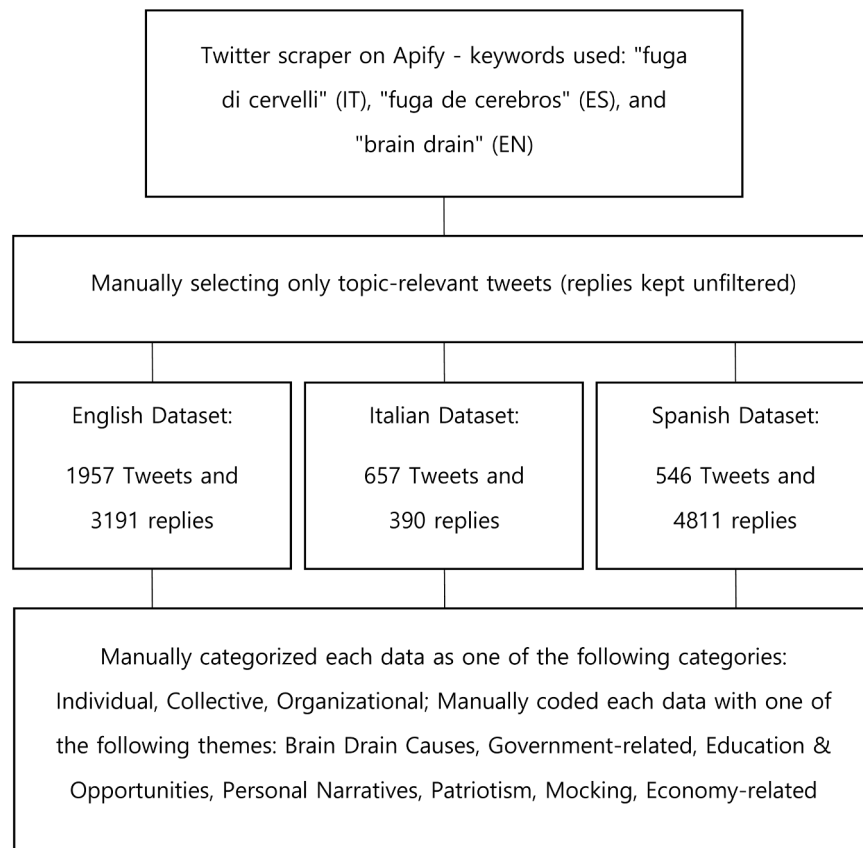
*Note.* Weights ( $\beta$  coefficients) are derived from language-specific OLS visibility models fitted separately for tweets and replies:  $\text{tweet\_retweets} \approx \beta_1 \times \text{tweet\_replies} + \beta_2 \times \text{tweet\_likes}$ ;  $\text{reply\_retweets} \approx \beta_3 \times \text{reply\_likes} + \beta_4 \times \text{reply\_replies}$ . All indices were z-standardized within language before modeling to ensure comparability of direction and relative magnitude, but not absolute scale, across languages.

**Table A3**  
Correlation between topic and thematic divergence.

Language	n Pairs	r (Pearson)	95 % CI	Spearman $\rho$
English	3132	0.12	[0.084, 0.156]	0.12
Spanish	4490	0.05	[0.018, 0.080]	0.03
Italian	253	0.31	[0.194, 0.429]	0.32

*Note.* All correlations computed on complete tweet–reply pairs; bootstrapped 95 % percentile confidence intervals ( $R = 1000$ ) shown in brackets. Moderate positive associations suggest conceptual alignment yet non-redundancy between topic-based and thematic divergence measures.

Fig. A1, Tables A1, A2, A3, A4, A5.



**Fig. A1.** Validation of extended Spanish sentiment lexicon.

*Note.* Scatterplot comparing baseline and extended lexicon sentiment scores (sent\_lex\_old vs. sent\_lex). Pearson  $r = -0.07$  (95 % CI [-0.10, -0.04]) indicates that adding migration-related terms, such as migración, exilio, fuga, refugiado, diáspora, xenophobia, introduced meaningful new lexical variance tied to migration discourse rather than random noise. The slight negative association reflects that these newly added terms are predominantly negative in valence within the context of “brain drain”.

**Table A4**  
Multicollinearity diagnostics across models and languages.

Term	Type	Value Range	Model(s)	Languages
Tweet_Visibility	VIF	1.02–1.34	3–5	English, Spanish, Italian
Reply_Visibility_Score	VIF	1.00–1.04	3–5	English, Spanish, Italian
Sentiment_Shift	VIF	1.01–2.24	3–5	English, Spanish, Italian
Tweet_Visibility × Sentiment_Shift	VIF	1.26–2.75	5	English, Spanish, Italian
Categories_replies	GVIF*(1/(2 × Df))	1.00–1.03	4–6	English, Spanish, Italian

*Note.* All VIF and GVIF values < 3.0, indicating no multicollinearity. Categorical predictors evaluated using the generalized VIF adjusted for degrees of freedom.

**Table A5**  
**HC3-robust regression estimates across models and languages.**

Estimated coefficients (with HC3 heteroskedasticity-robust standard errors) for Models 3–6 by language. Predictors include tweet-level and reply-level visibility indices, sentiment shift, and interaction terms. All key predictors retain significance and expected direction, confirming robustness to heteroskedasticity.

Language	Model	Term	Estimate	Std. Error	t-Statistic	p-Value
English	model3	Tweet_Visibility	1.001	0.031	32.13	< 0.001
		Reply_Visibility_Score	0.849	0.048	17.57	< 0.001
		Sentiment_Shift	3.77	20.87	0.18	.857
English	model4	Tweet_Visibility	1.009	0.031	32.42	< 0.001
		Reply_Visibility_Score	0.881	0.061	14.42	< 0.001
		Sentiment_Shift	-0.19	20.11	-0.01	.993
English	model5	Tweet_Visibility	0.983	0.038	25.78	< 0.001
		Reply_Visibility_Score	0.804	0.060	13.37	< 0.001
Spanish	model3	Tweet_Visibility	1.012	0.013	77.75	< 0.001
		Reply_Visibility_Score	0.983	0.327	3.00	.003
		Sentiment_Shift	130.68	231.52	0.56	.572
Spanish	model5	Tweet_Visibility	1.111	0.025	44.69	< 0.001
		Reply_Visibility_Score	0.983	0.324	3.03	.002

(continued on next page)

Table A5 (continued)

Language	Model	Term	Estimate	Std. Error	t-Statistic	p-Value
Italian	model3	Sentiment_Shift	−993.9	410.37	−2.42	.015
		Tweet_Visibility × Sentiment_Shift	2.704	0.469	5.77	< 0.001
		Tweet_Visibility	1.002	0.003	360.73	< 0.001
		Reply_Visibility_Score	1.123	2.234	0.50	.616
Italian	model4	Sentiment_Shift	16.68	4.90	3.40	< 0.001
		Tweet_Visibility	1.003	0.003	308.52	< 0.001
		Sentiment_Shift	16.75	4.91	3.41	< 0.001
Italian	model5	Tweet_Visibility	1.001	0.003	358.05	< 0.001
		Sentiment_Shift	17.10	5.47	3.12	.002

Note. Across all languages, coefficients for Tweet Visibility remain stable (~1.0) and highly significant. Reply Visibility and Sentiment Shift vary by language, with Italian showing a significant positive effect for sentiment. Interaction terms were generally nonsignificant, indicating additive rather than multiplicative effects. Across languages, a one-unit increase in standardized tweet visibility corresponds to roughly a one-unit increase in retweet impact, underscoring visibility as the dominant predictor. Spanish engagement also scales with visibility but shows more variance in reply effects. Italian sentiment coefficients are positive, supporting the role of affective resonance in smaller networks.

References

Ajoseh, S., Essien, E., & Ikhide, E. (2024). Brain drain and economic development: The case of sub-Saharan Africa. *Journal of African Development Studies*, 31(2), 145–162.

Benson, M. (2016). Deconstructing belonging in lifestyle migration: Tracking the emotional negotiations of the British in rural France. *European Journal of Cultural Studies*, 19(5), 465–481. <https://doi.org/10.1177/1367549416631547>

Bouchet-Valat, M. (2020). *SnowballC: Snowball stemmers based on the C libstemmer UTF-8 library* (R package version 0.7.0). <https://CRAN.R-project.org/package=SnowballC>.

Bourdieu, P. (1986). The forms of capital. In J. Richardson (Ed.), *Handbook of theory and research for the sociology of education* (pp. 241–258). Greenwood.

Brady, W. J., Crockett, M., & Van Bavel, J. J. (2021). The MAD model of moral contagion: The role of motivation, attention, and design in the spread of moralized content online. *Perspectives on Psychological Science*, 16(6), 1195–1218. <https://doi.org/10.1177/1745691620917336>

Carlino, C., Nolano, G., di Buono, M.P., & Monti, J. (2020). LaCulturaNonSiFerma—Report su uso e la diffusione degli hashtag delle istituzioni culturali italiane durante il periodo di lockdown. *arXiv preprint arXiv:2005.10527*.

Cau, E., Morini, V., & Rossetti, G. (2024). Trends and topics: Characterizing echo chambers’ topological stability and in-group attitudes. *PLOS Complex Systems*, 1(2). <https://doi.org/10.1371/journal.pcsy.0000008>

Charquero-Ballester, M., Walter, J. G., Rybner, A. S., Nissen, I. A., Enevoldsen, K. C., & Bechmann, A. (2024). Emotions on Twitter as crisis imprint in high-trust societies: Do ambient affiliations affect emotional expression during the pandemic? *PLOS ONE*, 19(3), Article e0296801. <https://doi.org/10.1371/journal.pone.0296801>

Chen, C., Bernard, A., Rylee, R., & Abel, G. (2022). Brain circulation: The educational profile of return migrants. *Population Research and Policy Review*, 41(1), 387–399. <https://doi.org/10.1007/s11113-021-09660-3>

Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology*, 94, S95–S120. <https://doi.org/10.1086/228943>

Docquier, F., & Rapoport, H. (2012). Globalization, brain drain, and development. *Journal of Economic Literature*, 50(3), 681–730. <https://doi.org/10.1257/jel.50.3.681>

Duong, H. T., & Nguyen-Thi, T. A. (2021). A review: Preprocessing techniques and data augmentation for sentiment analysis. *Computational Social Networks*, 8(1), 1. <https://doi.org/10.1186/s40649-021-00095-y>

Dubois, E., & Blank, G. (2018). The echo chamber is overstated: The moderating effect of political interest and diverse media. *Information, Communication & Society*, 21(5), 729–745. <https://doi.org/10.1080/1369118X.2018.1428656>

Estevez, M., Ballestar, M. T., & Sainz, J. (2025). Market research and knowledge using Generative AI: The power of large language models. *Journal of Innovation & Knowledge*, 10(5), Article 100796.

Genediazjr. (2017). *Stopwords list for multiple languages*. GitHub repository. Available at: <https://github.com/Genediazjr/stopwords>.

Gjerazi, B. (2024). Media literacy in the era of globalization: Innovative strategies for decoding information. *Interdisciplinary Journal of Research and Development*, 11(1S1), 18–18.

Hale, S. A., Gaffney, D., & Graham, M. (2022). Cross-language dynamics on Twitter. *Journal of Communication*, 72(4), 472–495. <https://doi.org/10.1093/joc/jqac002>

Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 168–177).

Hynek, N., Gavurova, B., & Kubak, M. (2025). Risks and benefits of artificial intelligence deepfakes: Systematic review and comparison of public attitudes in seven European countries. *Journal of Innovation & Knowledge*, 10(5), Article 100782.

Ibáñez-Sánchez, S., Flavián, M., Casaló, L. V., & Belanche, D. (2022). Influencers and brands successful collaborations: A mutual reinforcement to promote products and services on social media. *Journal of Marketing Communications*, 28(5), 469–486. <https://doi.org/10.1080/13527266.2020.1794483>

Karamouzas, D., Mademlis, L., & Pitas, I. (2022). Public opinion monitoring through collective semantic analysis of tweets. *Social Network Analysis and Mining*, 12(1), 91. <https://doi.org/10.1007/s13278-022-00936-4>

Kim, M., Jang, K., & Kim, H. (2025). Improving cross-lingual phonetic representation of low-resource languages through language similarity analysis. In *ICASSP 2025–2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1–5). IEEE. <https://doi.org/10.1109/ICASSP48485.2025.10012345>.

Nguyen, C. H. (2006). Brain drain and patriotism. *Social Sciences Information*, 45(3), 463–485. <https://doi.org/10.1177/0539018406066531>

Nielsen, F.Å. (2011). A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. *arXiv preprint, arXiv:1103.2903*. <https://arxiv.org/abs/1103.2903> Accessed: 10 November 2024

Novak, P. K., Smailović, J., Sluban, B., & Mozetič, I. (2015). Sentiment of emojis. *PLOS ONE*, 10(12), Article e0144296. <https://doi.org/10.1371/journal.pone.0144296>

Park, J. (2025). *Navigating the brain drain crisis: Visibility, sentiment, and divergence on X (formerly Twitter) among english, Spanish, and Italian users* [Data and code repository]. *Open Science Framework*. <https://doi.org/10.17605/OSF.IO/S6VZ2>

Patel, R., & Kim, J. (2025). Institutional trust and migration narratives in Southern Europe. *Journal of International Migration Studies*, 12(1), 55–78.

Polignano, M., Basile, P., de Gemmis, M., Semeraro, G., & Basile, V. (2019). Sentix: A sentiment lexicon for social media texts in Italian. In *Proceedings of the 6th Italian Conference on Computational Linguistics (CLIC-it 2019)*. CEUR.

Porfirio, J. A., Felício, J. A., Rodrigues, R. M., & Carrilho, T. (2024). Exploring migrant entrepreneurship and innovation in ultraperipheral regions: An investigation on opportunity and necessity-driven entrepreneurship. *Journal of Innovation & Knowledge*, 9(4), Article 100573.

Putnam, R. D. (2000). *Bowling alone: The collapse and revival of american community*. Simon & Schuster.

Rathje, S., Van Bavel, J. J., & Linden, S. (2021). Out-group animosity drives engagement on social media. *Proceedings of the National Academy of Sciences*, 118(26), Article e2024292118. <https://doi.org/10.1073/pnas.2024292118>

Rémy, F.C. (.2023). The impact of brain drain on Haiti’s rural communities: The case of a small town in Central Haiti and its surroundings [Unpublished manuscript].

Ruii, M. L., & Regnedda, M. (2022). Between online and offline solidarity: Lessons learned from the coronavirus outbreak in Italy. *American Behavioral Scientist*, 1–18. <https://doi.org/10.1177/00027642221132177>

Santoveña-Casal, S., Gil-Quintana, J., & Ramos, L. (2021). Digital citizens’ feelings in national #Covid 19 campaigns in Spain. *Heliyon*, 7. <https://doi.org/10.1016/j.heliyon.2021.e08112>

Segev, E. (2023). Sharing feelings and user engagement on Twitter: It’s all about me and you. *Social Media Society*, 1–13. <https://doi.org/10.1177/20563051231183430>. April-June 2023.

Tomić, C. H., & Taylor, K. (2018). The impact of brain drain on education and labor markets. *International Migration*, 56(2), 87–102. <https://doi.org/10.1111/imig.12400>

Vega-Muñoz, A., González-Gómez-del-Miño, P., & Espinosa-Cristia, J. F. (2021). Recognizing new trends in brain drain studies in the framework of global sustainability. *Sustainability*, 13(6), 3195. <https://doi.org/10.3390/su13063195>