

From Headlines to Policy: Uncovering Media Narratives on Immigration Through a Global Lens

Amit Kravchik

Ben-Gurion University of the Negev

Be'er Sheva, Israel

amitkra@post.bgu.ac.il

ABSTRACT

In an era where immigration continues to shape global politics and social discourse, understanding how news media frame these debates is essential. This paper provides a cross-national examination of newspaper coverage on immigration, drawing on data from the Global Database of Events, Language, and Tone (GDELT) for Canada, Germany, the United Kingdom, and the United States between 2018 and 2025. We combine a large-scale sentiment analysis with BERTopic modeling to identify the most salient themes and track how public crises—such as the COVID-19 pandemic or geopolitical conflicts—shift sentiment in the press. Our findings indicate that coverage is consistently negative across all four countries, with marked spikes in negativity linked to contentious policy debates, border enforcement issues, and social unrest (e.g., riots in the United Kingdom in 2024). Yet, certain crisis moments, notably those involving refugee admissions from conflict zones, tend to elicit more nuanced or empathetic coverage. This study underscores the influence of political polarization and security-oriented frames in shaping dominant immigration narratives. By highlighting the dynamic interplay between events and media discourse, we offer a deeper understanding of how news outlets can both inform and inflame public perceptions of immigration.

KEYWORDS

NLP, Immigration, Newspapers, Sentiment Analysis

ACM Reference Format:

Amit Kravchik. 2025. From Headlines to Policy: Uncovering Media Narratives on Immigration Through a Global Lens. In . ACM, New York, NY, USA, 6 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

1 INTRODUCTION

In the age of globalization, immigration remains one of the most divisive and discussed topics across the globe. Political and economic changes [2, 5], alongside global crises like the COVID-19 pandemic [1], have amplified debates about immigration policies and their socio-economic effects. Since media significantly influences public opinion and policy-making [6], examining the evolution of media tone and coverage is key to uncovering underlying biases.

The Global Database of Events, Language, and Tone (GDELT) [7] catalogs global media outputs in over 65 languages, offering a strong resource to explore how media narratives and sentiments about social topics are constructed and received globally. Using GDELT's datasets, researchers can conduct detailed comparative analyses of media narratives and sentiments, offering a broader

perspective on the cultural and political impacts of media coverage on immigration.

This study aims to utilize the comprehensive data capabilities of GDELT to explore several key aspects of media coverage on immigration:

- (1) **Trends Over Time:** To identify how the tone and extent of media coverage on immigration have evolved over recent years across different countries.
- (2) **Comparative Analysis:** To compare how media narratives and sentiments towards immigration vary between countries with different political and cultural contexts.
- (3) **Impact of Significant Events:** To examine how major global or national events influence media coverage and public discourse on immigration.

By achieving these objectives, the research seeks to contribute valuable insights into the role of media in shaping public discourse and policy on immigration, offering a data-driven foundation for further academic exploration and practical policy formulation.

The full implementation can be found at our public GitHub repository.¹

2 RELATED WORKS

Research on media coverage of immigration has consistently shown that the ways in which news outlets frame immigration issues can profoundly impact public opinion and policy outcomes. Early agenda-setting studies (McCombs & Shaw, 1972 [8]) have evolved into sophisticated analyses that examine not only the volume but also the content and tone of coverage.

The research conducted by McLaren, Boomgaarden, and Vliegenthart [9] highlights the powerful role of British newspapers in molding public perceptions of immigration. They discovered that media emphasis on the economic and security implications of immigration could significantly heighten public concern. This effect occurs despite the public's limited direct experience with these issues, underscoring the media's capacity to influence perceptions through selective coverage.

Lawlor's study [6] offers an in-depth comparison of how Canadian and British media frame immigration issues. Although the two countries have similar policies regarding immigration, their media narratives differ significantly. Lawlor found that variations in media framing are influenced by each country's distinct political, cultural, and institutional landscapes. This comparative approach underscores the importance of considering political, cultural, and institutional contexts when examining media effects on immigration attitudes.

¹<https://github.com/AmitKravchik/global-immigration-attitudes-analysis>

In their 2021 study, Young et al. observed that newspapers in the United States tend to concentrate on policies that exclude immigrants, especially during times of significant political division. Their analysis reveals that despite the development of policies at state and local levels aimed at integrating immigrants, the prevalent media narrative continues to portray immigrants negatively, often as criminals or economic liabilities. This persistent focus on exclusionary policies, they suggest, may adversely affect the health of immigrants and influence the direction of broader public policy debates.

Recent advances in computational social science have expanded the toolkit available for analyzing media content at scale. With the advent of large-scale datasets like GDELT, researchers can now examine cross-national media coverage in unprecedented detail. This computational approach complements traditional content analyses by providing a dynamic, longitudinal view of how immigration frames evolve in response to policy shifts and global events. Such methods have the potential to bridge gaps in our understanding by linking micro-level framing effects to macro-level trends in public opinion.

3 BACKGROUND

Selection of Target Countries. This study focuses on the United States, Germany, the United Kingdom, and Canada—countries that not only attract significant numbers of immigrants but also represent diverse approaches to immigration policy and public debate. The United States, for example, has traditionally been a primary destination for immigrants, where its dynamic and diverse population has significantly shaped its socio-economic landscape [10]. Germany, particularly after the surge of refugees in 2015, has witnessed significant policy shifts and heated debates regarding integration and national security [12]. Similarly, the United Kingdom has experienced intense public and political debate over immigration, a trend further amplified by its decision to leave the European Union, thereby reframing national discussions about borders and cultural identity [3]. Canada is widely recognized for its progressive and inclusive immigration policies, which have fostered a model of multiculturalism [11]. This makes it an ideal case study for examining how media coverage intersects with generally more positive public sentiment on immigration, despite sporadic debates over border management and resource allocation.

Together, these four countries provide a rich, diverse sample for exploring how media narratives on immigration vary across different socio-political and cultural landscapes.

Data Source and Format. Our study leverages data from the Global Database of Events, Language, and Tone (GDELT), a real-time open-access platform that monitors global news media in over 65 languages. GDELT continuously ingests and processes news articles from thousands of sources worldwide. Its automated systems not only translate non-English content but also analyze each article for key attributes—including thematic categorization, geolocation, and a composite sentiment “tone” score. The tone score, which ranges from -100 (extremely negative) to 100 (extremely positive), provides a quantitative measure of the emotional tenor of media coverage.

For our analysis, we used the GDELT Doc API in its ToneChart mode. The API returns a set of tone bins that represent distinct values of sentiment. Within each bin, the API provides:

- **Tone Bin:** The defined sentiment value.
- **Article Count:** The total number of articles that fall within that tone range.
- **Top Articles:** A list of the top 10 most relevant articles in that tone bin, as determined by GDELT’s ranking algorithm. The article’s details consist of its title and URL address.

This aggregated format enables us to capture macro-level trends in media sentiment over time without processing every single article individually.

4 DATA COLLECTION AND PROCESSING

4.1 Data Collection Methodology

Given the enormous volume of data processed by GDELT, we developed an efficient, scalable data collection pipeline in Python. Our pipeline was designed with a multi-stage architecture that supports concurrent processing using both **multithreading** and **multiprocessing**:

(1) API Querying:

We constructed multiple API queries to cover various combinations of parameters. Specifically, we issued queries for each target country over weekly intervals. Each query in ToneChart mode returns aggregated data (tone bins, article counts, and top articles) for a given week. To minimize latency, these queries were executed concurrently using a multithreaded approach.

(2) HTML Content Retrieval:

For each article returned by the API, we concurrently fetched the full HTML content using additional threaded HTTP requests. This step allowed us to obtain the complete text of each article, which was not fully available in the initial API response.

(3) Parsing and Processing:

The fetched HTML content was then processed using **BeautifulSoup** to parse and extract the main body and title of the articles. Since HTML parsing is computationally intensive, we implemented this stage using a multiprocessing approach. This allowed us to distribute the CPU-bound workload across multiple cores, bypassing the limitations of Python’s Global Interpreter Lock (GIL).

(4) Incremental Storage:

To manage memory usage and ensure scalability, our pipeline writes results to disk incrementally. As soon as an article has been fully processed, it is serialized to JSON and appended to a file using a dedicated multithreaded writer. This producer-consumer design ensures that data is persisted without waiting for the entire dataset to be collected in memory.

By combining multithreading for I/O-bound tasks (API querying and HTTP fetching) with multiprocessing for CPU-intensive processing, our pipeline efficiently collected and processed aggregated ToneChart data from GDELT. The resulting dataset, which consists of weekly tone bins with associated article counts and top articles,

forms the basis for our analysis of media sentiment trends related to immigration.

4.2 Data Summary

By querying GDELT’s ToneChart mode weekly for each country from January 2018 to January 2025, we initially retrieved a total of 170,914 articles across the United States, Canada, Germany, and the United Kingdom. Table 1 outlines the distribution of these articles by country, along with the proportion that contained valid HTML bodies.

Country	Articles Collected	Valid HTML (%)
United States	56,346	95.5
United Kingdom	44,304	99.76
Germany	31,011	98
Canada	39,253	99.92

Table 1: Dataset Statistics by Country (2018–2025)

From this total, 98% of the articles contained valid HTML content and passed our initial quality checks. For the remaining documents, we use the title as the content. In the next section, we detail how we processed and aggregated these articles for sentiment analysis, topic modeling, and entity extraction.

Figure 1: Articles tone over time.

5 ANALYSIS

5.1 Overall Trends in Coverage and Tone

Across all four countries—Canada, Germany, the United Kingdom, and the United States—media discourse on immigration exhibits a consistently negative baseline tone. As seen in Figure 2, the mean tone for each country rarely approaches neutrality, indicating a predominantly critical or alarmist framing of immigration issues. Despite occasional months where sentiment becomes relatively less negative, there is no sustained shift toward more neutral or positive coverage in any of the observed time spans.

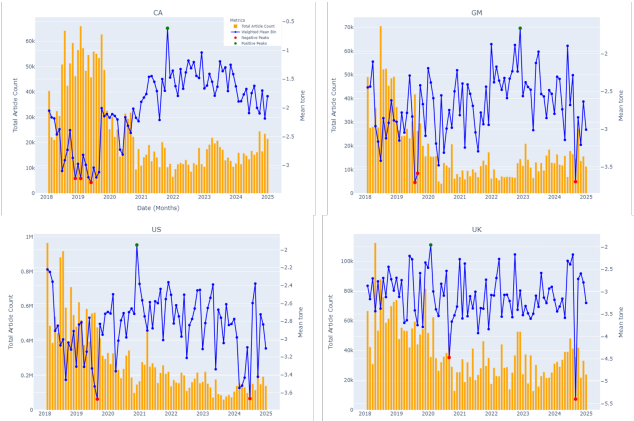


Figure 2: Monthly media coverage and tone across countries, highlighting periods of extreme sentiment through peaks in weighted mean tones.

Coverage Volume. One striking difference among the four countries lies in the volume of articles published. The United States leads in sheer coverage, often surpassing several hundred thousand articles monthly and occasionally reaching close to one million. This heightened media attention could reflect a combination of the country’s extensive media infrastructure and a high-profile national debate on immigration policies. In contrast, the United Kingdom and Germany typically range between tens of thousands to over one hundred thousand articles per month.

Notably, both countries experience spikes at various points in time—particularly in 2019 and 2021—suggesting major political or social events that reignited immigration discourse. However, another explanation worth considering is the potential for anomalies or inconsistencies in the data collection methods used by GDELT, which could have led to an overrepresentation during this period.

Tone Trajectory. In addition to coverage volume, the mean tone remains consistently negative for each country, though the degree of negativity differs. Germany and the United States frequently dip below -3.0 , suggesting an especially critical or negative framing. Canada’s tone mostly hovers between -1.5 and -2.5 , while the United Kingdom remains around -2.0 to -2.5 . Despite these variations, the general conclusion is that no country experiences an extended shift toward neutral or positive sentiment during the 2018–2025 period (see Figure 1).

Short-term peaks do appear, as marked by red (negative) or green (positive) markers in the charts. These indicate months where tone diverges significantly from the average—either because of a brief uptick (less negative) or a sharp downturn (more negative). Such short-term spikes often coincide with surges in coverage, implying that heightened media attention may exacerbate or intensify the critical tone.

Comparative Observations. When viewed comparatively, Canada’s early high coverage and moderately negative tone gradually give way to a more stable, though still critical, pattern in later years. Germany and the United Kingdom each show moderate-to-high coverage volumes with deep negativity that fluctuates but does

not trend upward toward neutrality. The United States consistently demonstrates both the highest coverage volume and one of the most negative tones, occasionally reaching less negative levels in certain months before reverting to more critical framing. These cross-country contrasts point to potentially different socio-political drivers of immigration discourse, as well as distinct media environments that shape how stories are told and which aspects of immigration gain traction.

These overall trends raise questions about the specific events, narratives, and policy debates that drive surges in coverage and dips in tone. The marked negativity across all countries underscores the need for deeper inquiry into the content of the articles, including which topics dominate discussions and which entities (politicians, organizations, geographic locations) are most frequently mentioned. The subsequent sections of this paper will delve into these details using topic modeling and entity extraction, focusing particularly on periods where sentiment diverges most sharply from the average.

5.2 Analysis of Extreme Sentiment Peaks

Before examining the specific events that triggered extreme sentiment shifts, we define a *peak* as any monthly tone value that deviates by more than two standard deviations from the overall mean. These peaks represent moments of particularly intense media framing—either strongly negative or moderately positive—when immigration coverage becomes exceptionally salient.

To identify the dominant themes in these peak periods, we employed **BERTopic** [4], a state-of-the-art topic modeling library that uses transformer-based embeddings to group semantically similar documents. We then extracted the most salient named entities from these same periods using **spaCy**, a widely used natural language processing toolkit known for its robust entity recognition capabilities. The synergy of these methods—topic modeling and entity extraction—helps us pinpoint not only *what* was being discussed but also *who* and *where* were central to these heightened spikes in sentiment.

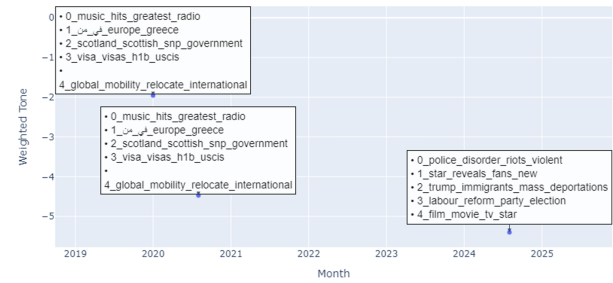
Integrating the **BERTopic** modeling results with the sentiment analysis reveals distinct periods of heightened media attention and intensified negative or positive coverage across the four countries examined. These peaks—whether strongly negative or moderately positive—are closely tied to real-world political developments, humanitarian crises, or social unrest that brought immigration to the forefront of public discourse.

Temporal Alignment with Key Events. A clear pattern emerges when overlaying the temporal frequency of BERTopic clusters onto sentiment trajectories. Notable dips in sentiment coincide with major policy debates, border controversies, and large-scale migrations (see Figures 3,4):

- **United States:**

- **Late 2018 to early 2019:** Articles clustered under “family separations” and “asylum restrictions” show a sharp rise in volume, reflecting extensive coverage of migrant families being held in detention centers. Negative sentiment peaks here, aligning with heated public and congressional debates over child welfare and border control.

Peak Events: Tone Over Time for UK



Peak Events: Tone Over Time for US

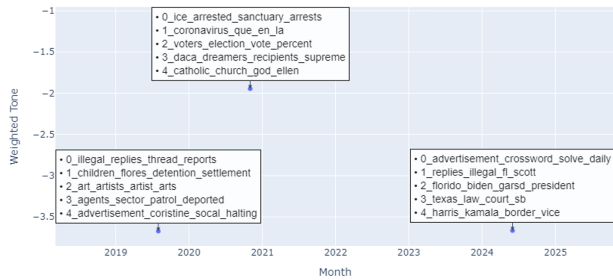


Figure 3: Monthly media coverage and tone across countries, highlighting periods of extreme sentiment through peaks in weighted mean tones.

- **2023–2024:** Another negative spike is evident in clusters relating to “state vs. federal immigration laws,” with coverage often framing immigration as a crisis or emphasizing legislative conflicts in states like Florida.

- **Canada:**

- **Late 2018–Early 2019:** Clusters around “irregular border crossings” into Quebec surge. Local media raise concerns over resource allocation and social cohesion, driving an uptick in negative sentiment.
- **2021–2022:** While overall coverage becomes less intense, topics such as “Afghan refugee resettlement” and “COVID-related travel restrictions” maintain a mildly negative tone, partly due to bureaucratic backlogs and pandemic-related uncertainties.

- **Germany:**

- **2019–2020:** Marked by a pronounced trough in sentiment corresponding to topic clusters on “EU refugee quotas” and “domestic integration challenges.” Lingering tensions from the 2015–2016 influx continued to polarize debates about responsibility-sharing within the EU.
- **2022–2023:** Initially, the Russian invasion of Ukraine prompted more empathetic coverage of Ukrainian refugees. However, criticism regarding capacity limits and unequal treatment of non-European refugees led to intermittent negative sentiment peaks.

- **United Kingdom:**

- **2019–2020:** Clusters focused on “Brexit negotiations” and “ending free movement” reveal a distinct negative sentiment peak. Intense media scrutiny, especially from tabloid outlets, amplified concerns about labor market impacts and sovereignty.
- **2024–2025:** Another notable negative trough appears in the data, associated with clusters addressing “mass deportations” and “tougher enforcement measures.” Media coverage of **riots** in several urban centers escalated hostility toward immigration policies, further lowering sentiment.

Dominant Framing of Immigration. Close examination of the BERTopic clusters indicates that newspapers frequently employ either security-oriented, economic, or humanitarian frames. Whenever border controversies surface—for instance, debates on a potential border wall in the U.S. or irregular entries in Canada—tone typically skews negative and uses crisis-oriented language. Conversely, humanitarian frames come to the fore during refugee crises (e.g., with Afghan or Ukrainian arrivals), occasionally resulting in more empathetic tones, though this empathy is often tempered by discussions about resource constraints or public backlash.

Political Polarization and Policy Debates. One of the most conspicuous drivers of negative sentiment spikes is political polarization. In the United States, partisan disputes over enforcement actions, sanctuary cities, and DACA frequently elevate immigration as a flashpoint issue in the press. In Germany, divisions between centrist and right-wing parties on refugee admission policies add to negative coverage whenever budgetary or cultural integration questions arise. Meanwhile, in the U.K., contentious Brexit negotiations and subsequent reforms, including debates over points-based systems and visa restrictions, consistently drive negativity. The 2024 riots further illustrate how public protest and media sensationalism intensify hostility toward immigration policies, as evidenced by the heavily negative framing in certain tabloid outlets.

Overall, the integrated BERTopic and sentiment analysis underscores the responsiveness of newspaper coverage to political, economic, and humanitarian triggers. Each distinct drop or surge in sentiment can be traced to an uptick in coverage around particular topics—ranging from family separation policies in the U.S. to irregular border crossings in Canada, from EU quota debates in Germany to Brexit-related reforms and riots in the U.K. This alignment suggests that, far from being random fluctuations, media sentiment peaks appear when high-stakes policy decisions or sudden crises force immigration issues to the center of public debate. In these periods, framing tends to polarize quickly, pushing coverage to a predominantly negative tone, with occasional counterpoints when humanitarian narratives gain temporary traction.

5.2.1 Entity Extraction in Peak Events. Alongside topic modeling and sentiment analysis, we performed entity extraction to pinpoint the most frequently mentioned individuals, organizations, and locations at each sentiment peak.

- **Canada (Late 2018–Early 2019):** Mentions of “Trump,” “Quebec,” “Mexico,” and “Canadian” dominate, reflecting coverage of U.S. policy shifts, irregular border crossings, and local resource debates.

- **Germany (Mid–Late 2019):** Entities include “Wolfsburg,” “Salzgitter,” and “Peine” (cities in Lower Saxony), alongside “Trump” and “USA,” indicating a mix of local integration concerns and international references to U.S. immigration policies.
- **United Kingdom (2020 and beyond):** Clusters mentioning “UK,” “French,” “Dover,” “Channel,” and occasionally “Trump” or “Harry” show a dual focus on cross-channel migration issues and high-profile figures shaping immigration debates.
- **United States (2019–2024):** Political figures (“Trump,” “Biden”) and regions (“Texas,” “Mexico”) frequently appear. These peaks correspond to heightened coverage of border enforcement, election-driven policy shifts, and local contexts (e.g., Texas legislation).

Overall, entity extraction clarifies which actors and places drive coverage spikes. During highly negative peaks, news outlets often highlight political leaders, contested borders, or specific localities, reinforcing polarized discussions about immigration. By contrast, the presence of broader or more general terms (e.g., “America,” “Britain”) sometimes signals a shift toward cultural or societal narratives, showing how media attention expands or narrows depending on the issue at hand.

6 CONCLUSION

This research demonstrates that media coverage of immigration follows a cyclical pattern driven by political, economic, and humanitarian events. Across the four countries—Canada, Germany, the United Kingdom, and the United States—negative sentiment generally persists but intensifies when high-stakes events intersect with critical policy debates. The BERTopic analysis reveals that newspapers selectively amplify certain aspects of immigration based on real-world triggers and editorial priorities.

These findings imply that media narratives do more than reflect policy debates; they can also shape the urgency and tone of those debates. Although humanitarian frames sometimes yield more empathetic portrayals, they rarely produce sustained improvements in overall sentiment. Future studies should expand geographic scope and integrate longitudinal polling data to further clarify the relationship between media depictions and public attitudes.

7 ACKNOWLEDGMENTS

This work was carried out using data from the Global Database of Events, Language, and Tone (GDELT). In addition, we employed several Python libraries, including **spaCy** for entity extraction, **pandas** and **NumPy** for data handling, **scikit-learn** for various machine learning utilities, and **BERTopic** for topic modeling. Portions of this manuscript were composed with the assistance of **ChatGPT**, a large language model developed by OpenAI, which provided editing and phrasing suggestions. We extend our gratitude to **Dr. Michael Fire**, whose guidance and support as a course lecturer played a significant role in shaping this project.

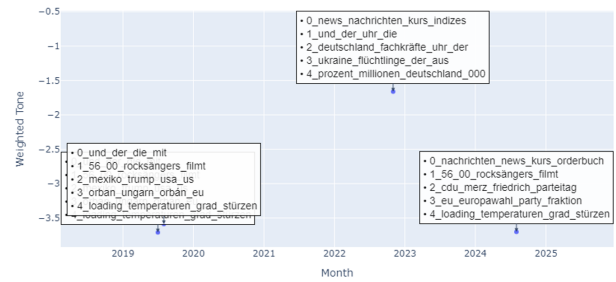
REFERENCES

- [1] Helen Canton. International organization for migration—iom. In *The Europa Directory of International Organizations 2021*, pages 350–354. Routledge, 2021.
- [2] Christian Dustmann, Tommaso Frattini, and Caroline Halls. Assessing the fiscal costs and benefits of a8 migration to the uk. *Fiscal Studies*, 31(1):1–41, 2010.

- [3] Matthew Goodwin and Caitlin Milazzo. Taking back control? investigating the role of immigration in the 2016 vote for brexit. *The British Journal of Politics and International Relations*, 19(3):450–464, 2017.
- [4] Maarten Grootendorst. Bertopic: Neural topic modeling with a class-based tf-idf procedure. *arXiv preprint arXiv:2203.05794*, 2022.
- [5] Jens Hainmueller and Michael J Hiscox. Educated preferences: Explaining attitudes toward immigration in europe. *International organization*, 61(2):399–442, 2007.
- [6] Andrea Lawlor. Framing immigration in the canadian and british news media. *Canadian Journal of Political Science/Revue canadienne de science politique*, 48(2): 329–355, 2015.
- [7] Kalev Leetaru and Philip A. Schrodt. Gdelt: Global data on events, location, and tone. *ISA Annual Convention*, 2013. URL <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.686.6605>.
- [8] Maxwell McCombs and Donald Shaw. The agenda-setting function of mass media. *The Agenda Setting Journal*, 1, 09 2017. doi: 10.1075/asj.1.2.02mcc.
- [9] Lauren McLaren, Hajo Boomgaarden, and Rens Vliegthart. News coverage and public concern about immigration in britain. *International Journal of Public Opinion Research*, 30(2):173–193, 2018.
- [10] Alejandro Portes and Rubén G Rumbaut. *Immigrant America: a portrait*. Univ of California Press, 2006.
- [11] Jeffrey G Reitz. The distinctiveness of canadian immigration experience. *Patterns of prejudice*, 46(5):518–538, 2012.
- [12] Karen Schönwälder and Triadafilos Triadafilopoulos. The new differentialism: Responses to immigrant diversity in germany. *German Politics*, 25(3):366–380, 2016.
- [13] MARIA-ELENA DE TRINIDAD Young, Hannah Sarnoff, Dana Lang, and A Susana Ramirez. Coverage and framing of immigration policy in us newspapers. *The Milbank Quarterly*, 100(1):78–101, 2022.

8 APPENDIX

Peak Events: Tone Over Time for GM



Peak Events: Tone Over Time for CA

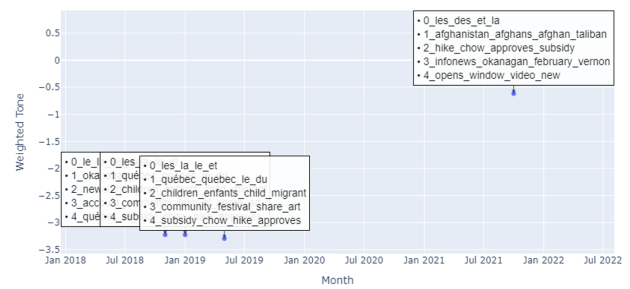


Figure 4: Monthly media coverage and tone across countries, highlighting periods of extreme sentiment through peaks in weighted mean tones.