News headline analysis of the 2024 US election

Minh Nguyen

2025-03-01

Table of Contents

1 Introduction

The political system of the United States is mainly structured around two major parties: the Democratic Party and the Republican Party. As a global superpower, the country's policies and leadership changes can have a significant impact on the world. This makes the presidential election one of the most important political events, held every four years. During the election process, candidates discuss topics of national and international importance, such as ongoing armed conflicts, economic policies, gun control legislation, and abortion rights. At the same time, a substantial majority (86%) of U.S. adults report accessing news via smartphones, computers, or tablets at least occasionally [1]. Additionally, the Harvard University's Future of Media Project has documented 176 parent companies and standalone news outlets within mainstream media [2]. With this level of media consumption, one may ask whether the media can influence how people vote—and if so, to what extent? This paper therefore analyzes politically oriented headlines related to the 2024 U.S. election to investigate whether there is a systematic way of presenting topics that could influence public opinion.

1.1 Context of the 2024 US Election

The 2024 U.S. presidential election occurred on November 5, 2024, under a tense political atmosphere of economic challenges and international tensions. This election is a rematch between former President Donald Trump and Vice President Kamala Harris, following President Joe Biden's withdrawal from the race in July 2024 due to health concerns [3] [4]. Donald Trump, representing the Republican Party, selected Senator J.D. Vance of Ohio as his running mate. The Republican focused on economic development, tight immigration control, and a robust national defense. On the other side, Kamala Harris, representing the Democratic Party, chose Governor Tim Walz of Minnesota as her vice-presidential candidate. The Democratic campaign prioritized social justice, healthcare reform, and the protection of civil rights [5].

The U.S. presidential election is also a complicated process, featuring several critical stages:

- Primaries and Caucuses: From February to June 2024, these events determined the delegates pledged to various candidates.
- Party Conventions: The Democratic National Convention took place from August 19–22, 2024, in Chicago, Illinois, where Kamala Harris was officially nominated. The Republican National Convention was held earlier, finalizing Trump's nomination [6].
- General Election Campaign: Post-conventions, both parties engaged in nationwide campaigns, participating in the general election on November 5, 2024.
- Electoral College Vote: Electors met in their respective state capitals on December 17, 2024 to cast their votes.
- Inauguration: The President-elect is scheduled to be inaugurated on January 20, 2025, marking the commencement of the new presidential term.

The election resulted in Donald Trump narrowly defeating Kamala Harris, with pivotal swing states such as North Carolina, Georgia, and Pennsylvania playing crucial roles in the outcome [5]. The US Sun Specifically, Donald Trump secured 312 electoral votes, surpassing the required 270, thereby winning the presidency.

2 Hypothesis

Media bias and framing has been used in various fields, including political science. McCombs and Shaw showed that news coverage is not merely a summary of information but can be a way to influences pulic opinion [7]. Additionally, the theory of framing suggests that the way information is organized can shape how audiences interpret political issues [8]. In the context of the 2024 U.S. presidential election, it is expected that news coverage, particularly through headlines, will differ systematically between left-leaning and right-leaning sources. Thus, the general hypothesis would be there is a difference in multiple aspects of the headlines between left- and right-leaning sources, that contribute to the public opinion and voting behaviors. [1] The aspects can be defined as follow:

Firstly, left-leaning headlines may feature more positive and bright tones, while right-leaning ones may have more aggressive and negative emotions, specially near election days [9] [10] [11] [12]. Therefore, this study hypothesizes that right-leaning outlets will display higher levels of negative sentiment, while left-leaning sources may emphasize positive emotions or hope-oriented themes to encourage their audiences.

A second difference is the presentation of identical topics. For instance, coverage on immigration might be framed as a humanitarian issue in left-leaning outlets, using phrases such as "asylum seekers in need of protection". Meanwhile, right-leaning outlets may frame it as a security threat, emphasizing terms like "border crisis". As multiple studies have suggested that the linguistic choices in headline phrasing can influence public perception [13] [14], this study expects to find systematic differences in lexical choices, keyword associations, and bi-gram structures between two leaning sources when covering common election topics.

Finally, framing theory suggests that left-leaning and right-leaning media prioritize different topics [14]. Specifically, left-leaning outlets are expected to focus more on social justice issues such as abortion rights, healthcare, and civil liberties, whereas right-leaning media are likely to emphasize economic performance, border security, and law enforcement policies. This hypothesis builds on Stroud's selective exposure theory, which suggests that media consumers gravitate toward news sources that reinforce their pre-existing beliefs [15].

[1] The counter-hypothesis (H0) would suggest that no significant differences exist in the sentiment trends, framing strategies, or topics between left-leaning and right-leaning news outlets, implying that media bias does not significantly alter election coverage.

3 Methods

3.1 Dataset

The dataset used in this study was obtained from Media Cloud, a platform that aggregates news articles and headlines from a wide range of media sources. To examine media bias and framing in the coverage of the 2024 U.S. presidential election, a diverse selection of news outlets was chosen, representing different political orientations. Based on domain knowledge from [16] [17] [18] , the media outlets included in this study are classified as follows:

- Left-leaning outlets: The New York Times (nytimes.com), the Nation (thenations.com), CNN (cnn.com), Los Angeles Times (latimes.com).
- Right-leaning outlets: Fox News (foxnews.com), New York Post (nypost.com), Breitbart (breitbart.com), Hot Air (hotair.com).

The dataset spans the period from January 15, 2024, to November 5, 2024, covering the entire election cycle, from the early primaries to election day. This ensures all the major electoral developments and campaign strategies were included in the analysis. In the end, there was about 3940 headlines relating to 2024 US election for left-leaning outlets, 3686 headlines for right-leaning outlets, which is suitable for the hypothesis involving comparisons between the two.

3.2 Data preprocessing

Before conducting sentiment analysis and topic modeling, all headlines underwent a series of text preprocessing steps to prepare the data for reliable and meaningful natural language analysis. These steps are standard in Natural Language Processing (NLP) workflows and aim to reduce noise, improve consistency, and ensure that the analysis reflects semantic content rather than superficial textual variation. The first step was tokenization, in which each headline was broken down into individual tokens (words or punctuation marks). To obtain the most relevant with high semantic tokens, all the text was converted to lowercase and stop words (such as “the”, “is”, “and”, “to”) were removed using nltk library [19]. By doing so, the plot below was obtained, showing the words and their frequency in the headlines.

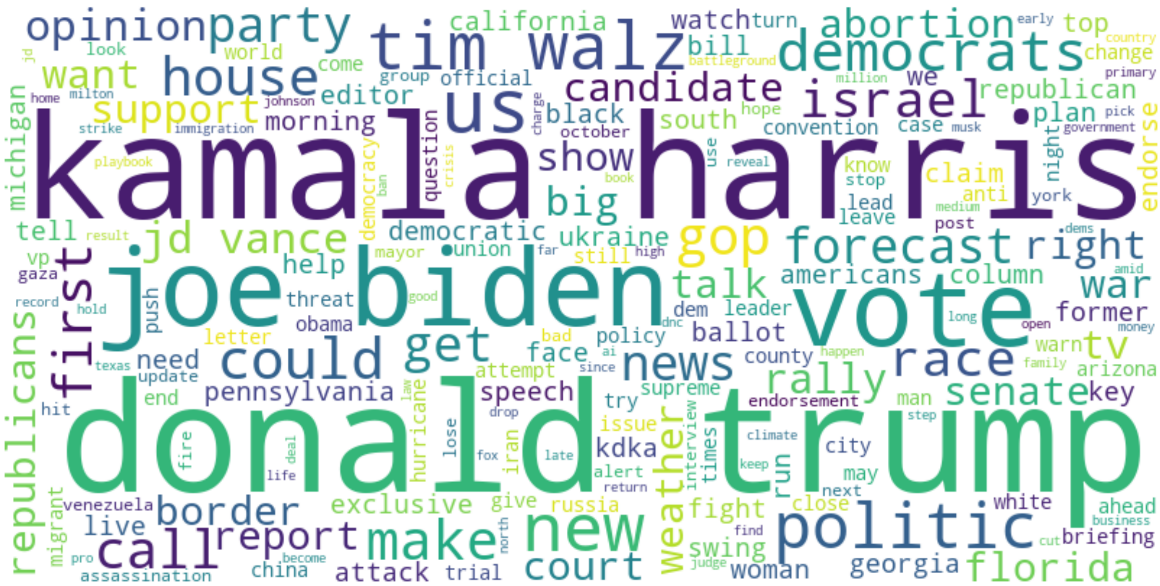

Figure 1: Wordcloud for US Election 2024.

From the plot, it can be observed that there is various trivial domain-specific words (“vote”,”show”,”news”) and names of political figures that might dominate headlines without contributing to framing analysis (e.g., “Trump,” “Biden,” “Kamala”). Therefore, a custom list of stop words was also added to remove those words, improving the quality of TF-IDF analys in the following sections. Lemmatization, using WordNetLemmatizer, was also applied to reduce word to their base forms (e.g., “voting” → “vote,” “immigrants” → “immigrant”) [20]. This allows grouping semantically identical terms under a single form, improving the accuracy for the following analysis. Finally, some similar words that cannot be lemmatized were also modified by synonym mapping. For example, “migrant”, “immigrant” were changed to “immigration”, “lgbtq”, “gay” were changed to “gender”. These techniques enhance the consistency for processing of some topics that features different rare words in the headlines

3.3 Sentiment analysis

Sentiment analysis is a Natural Language Processing (NLP) technique that determines the emotional polarity of text—whether it is positive, negative, or neutral—and, in some cases, the specific emotions expressed. It operates by mapping words in a text to predefined sentiment lexicons, which assign sentiment values based on prior human annotations.

3.3.1 NRC Emotion Lexicon

The NRC Emotion Lexicon is a word-based sentiment and emotion classification tool that maps words to eight basic emotions (joy, anger, fear, sadness, disgust, surprise, anticipation, trust) and two sentiment categories (positive, negative) [21]. Each word in a given text is matched against the lexicon, and the frequency of words associated with each category is counted. This method enables multi-label classification at the headline level, meaning a single headline can express multiple emotions simultaneously (e.g., a headline about a terrorist attack may score high for both *fear* and *anger*). Not only restricting to headlines, this approach can provide lists of words with desired emotions for deeper investigations. Mathematically, given a set of words $W = \{w_1, w_2, ..., w_n\}$ in a headline, the emotion score for an emotion category E_i is computed as:

$$S(E_i) = \sum_{j=1}^{n} I(w_j, E_i)$$

where $I(w_j, E_i)$ is an indicator function that returns 1 if word w_j is associated with emotion E_i in the NRC lexicon, and 0 otherwise. The variable n is the total number of words in the document.

3.3.2 AFINN Sentiment Scoring

The AFINN lexicon is a continuous sentiment scoring system, where words are assigned numeric sentiment values ranging from -5 (most negative) to +5 (most positive) [22]. A positive sentiment score indicates a predominantly positive sentiment, while a negative sentiment score reflects a more negative tone. This approach provides a more quantifable analys for the emotion. Formally, the sentiment score for a headline is computed as:

$$S = \sum_{j=1}^{n} s(w_j)$$

where $s(w_j)$ is the sentiment score of word w_j in the AFINN lexicon, and n is the total number of words in the document. This method would be then combined with Bigram analys, to give comparable values of emotions between the Bigrams.

3.4 Keyword analysis

3.4.1 Term Frequency-Inverse Document Frequency (TF-IDF)

TF-IDF is a statistical measure used to evaluate how important a word is in a document relative to a collection of documents (the corpus) [23]. It helps identify words that are significant within individual articles while discounting commonly used terms. The TF-IDF score for a word w in document d is given by:

$$\text{TF-IDF}(w, d) = \text{TF}(w, d) \times \text{IDF}(w)$$

Where:

- **Term Frequency (TF)** measures how frequently a word appears in a document.

$$\text{TF}(w, d) = \frac{f_{w,d}}{\sum_{w' \in d} f_{w',d}}$$

  where $f_{w,d}$ is the number of times word w appears in document d, and the denominator is the total word count in d.

- **Inverse Document Frequency (IDF)** reduces the weight of words that appear too frequently across many documents.

$$\text{IDF}(w) = \log \left(\frac{N}{1 + |D_w|} \right)$$

  where N is the total number of documents, and $|D_w|$ is the number of documents containing word w. The addition of 1 in the denominator prevents division by zero.

3.4.2 Word Correlation Analysis

Beyond simple frequency analysis, we examine word correlations, which reveal how words tend to co-occur within headlines. Correlation measures the linear relationship between the frequency of two words across documents [24]. A high positive correlation indicates that the two words often appear together, while a negative correlation suggests they rarely co-occur. The correlation coefficient between two words x and y is calculated as:

$$\rho(x, y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2} \sqrt{\sum (y_i - \bar{y})^2}}$$

where:

- x_i and y_i are the term frequencies of words x and y in document i,
- $\bar{x}$ and $\bar{y}$ are their respective mean frequencies across all documents.

3.4.3 Bigram Extraction

To better understand how key terms are framed, bigrams were also used, which are sequences of two consecutive words. Extracting bigrams allows detecting common phrase structures and how certain keywords are contextualized [25]. Formally, given a sequence of words $w_1, w_2, ..., w_n$, the bigram model computes:

$$P(w_i \mid w_{i-1}) = \frac{\text{Count}(w_{i-1}, w_i)}{\text{Count}(w_{i-1})}$$

where:

- $\text{Count}(w_{i-1}, w_i)$ is the number of times the bigram (w_{i-1}, w_i) appears in the corpus,
- $\text{Count}(w_{i-1})$ is the total number of occurrences of the first word in the bigram.

3.5 Topic modeling

Lastly, to uncover the underlying themese in the headlines, Latent Dirichlet Allocation (LDA) was applied on the headlines. LDA is a unsupverised probabilistic model that gives maps the topics obtained from the texts in the headlines to the actual headlines [26]. It works by assuming that each document (in this case, a headline) is a mixture of various latent topics, and each topic is characterized by a distribution over words. It aims to reverse-engineer the hidden structure of the corpus by inferring two distributions: the distribution of topics for each document and the distribution of words for each topic.

Formally, let:

- D: the number of documents (headlines)
- K: the number of latent topics
- V: the vocabulary size
- $w_{d,n}$: the n-th word in document d

LDA models the generative process of each document as follows:

1. For each topic $k \in \{1, ..., K\}$, draw a distribution over words:

$$\phi_k \sim \text{Dirichlet}(\beta)$$

2. For each document $d \in \{1, ..., D\}$, draw a distribution over topics:

$$\theta_d \sim \text{Dirichlet}(\alpha)$$

3. For each word $w_{d,n}$ in document d:

    1. Choose a topic $z_{d,n} \sim \text{Multinomial}(\theta_d)$

    2. Choose a word $w_{d,n} \sim \text{Multinomial}(\phi_{z_{d,n}})$

Here, α and β are hyperparameters controlling the sparsity of the document-topic and topic-word distributions, respectively. The output of the LDA algorithm is a set of topics, each represented by a ranked list of words most strongly associated with that topic, and a topic distribution for each document.

4 Results

4.1 Sentiment Analysis

Figure 2 below shows a similarity in sentiments of headlines from Left-leaning and Right-leaning outlets. It can be observed that the sentiment distributions are quite similar between the two leaning, with "negative" being the most common emotion.

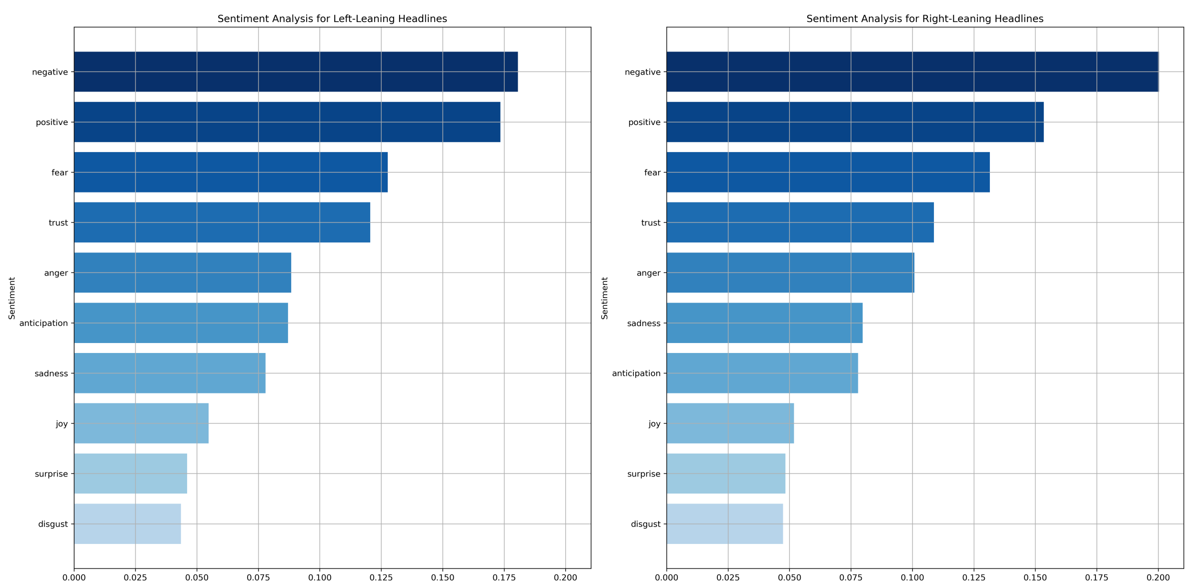

Figure 2: Sentiments' normalized scores for Left-leaning and Right-leaning outlets

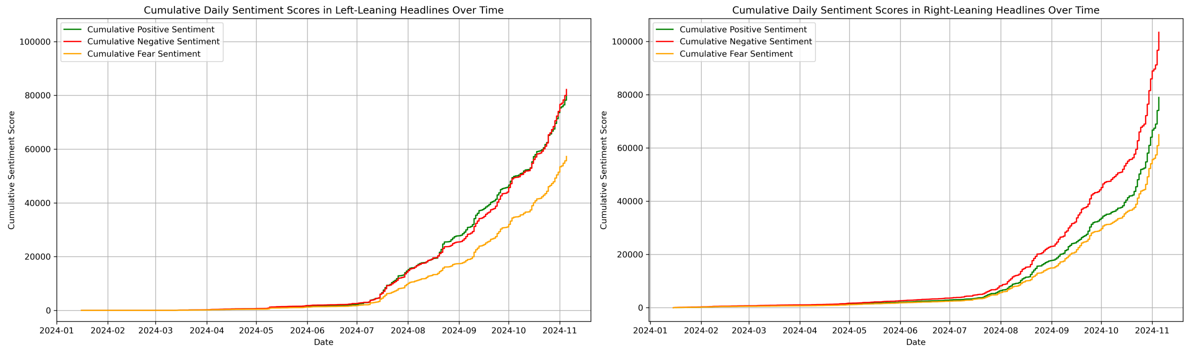

Figure 3: Sentiment trend over time

However, when plotting the sentiment over the whole election period, as shown in Figure 3, the positive and negative sentiments tend to be similar, with positive even took the lead at some points for the left-leaning outlents. On the other hand, the negative emotion dominants in the last four months. Additionally, fear emotion was almost the same as the positive emotion in right-leaning, showing some characteristics that were expected from the hypothesis section. This result might not support the counter-hypothesis to be rejected yet, but a difference between the two leaning sources could be observed. Using sentiment analysis, one can get a grasp of the topical differences between the leanings as well, as shown in Figure 4 and 5 below.

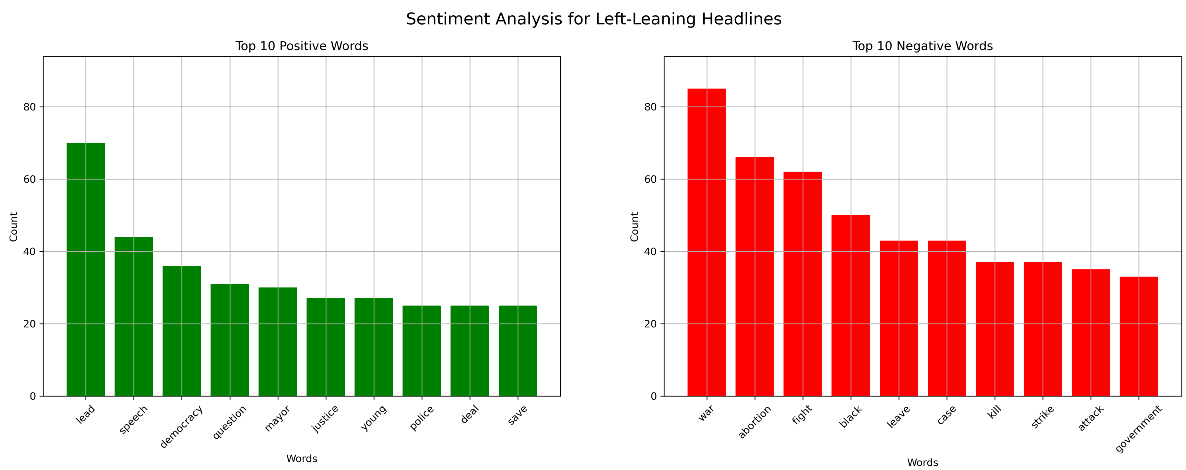

Figure 4: Positive and negative words of left-leaning outlets

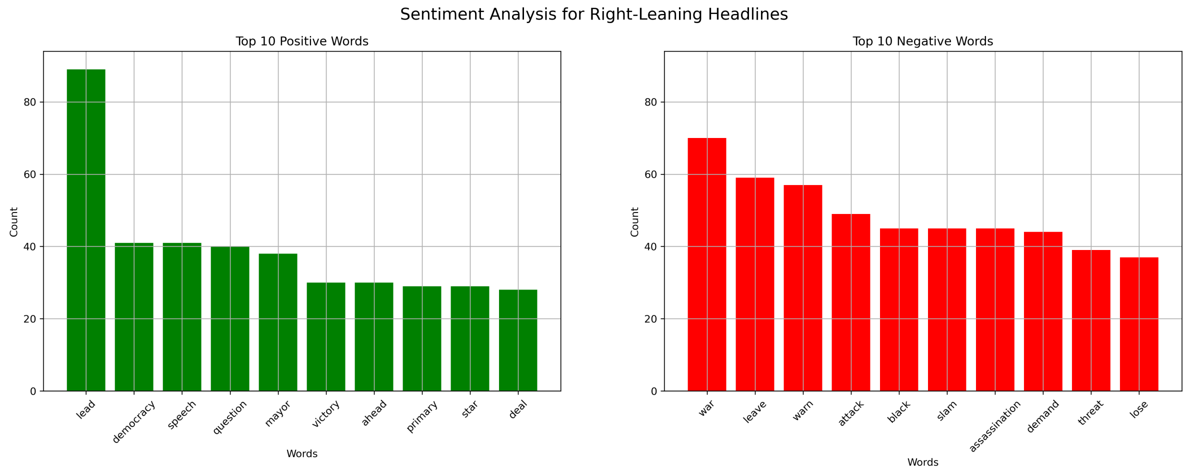

Figure 5: Positive and negative words of right-leaning outlets

From the plot, the positive words tend to appear quite similar between the two leanings. Regarding negative words, while the two leanings have some words in common (“war”, “black”, “attack”), the left-leaning headlines also include “abortion” and “strike”, reflecting its social justice characteristics as disccused in the hypothesis section. The common words would be studied further in the following sections.

4.2 Keyword analysis

Using the tf-idf metrics, the most important words for each leaning can be shown in Figures 6 and 7 below. The pivotal words for both leanings are “win” and “israel”, indicating headlines regarding the election results and the Israel-Palestine situation. Additionally, as expected, the left-leaning outlets considered words such as “woman” and “abortion” to be more important, as they are indicating social topics. On the other hand, the right-leaning side considered words like “immigration”, “border”, and “china” - words relating to border control and economy policies - more important.

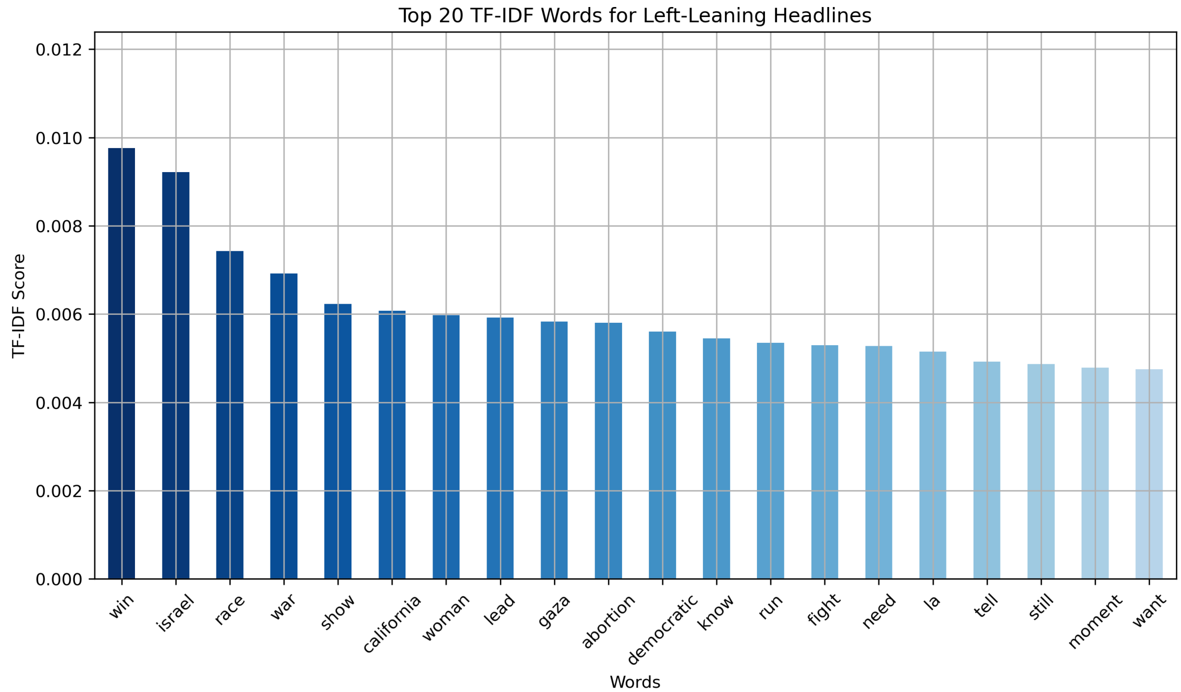

Figure 6: Tf-idf ranking of left-leaning headlines

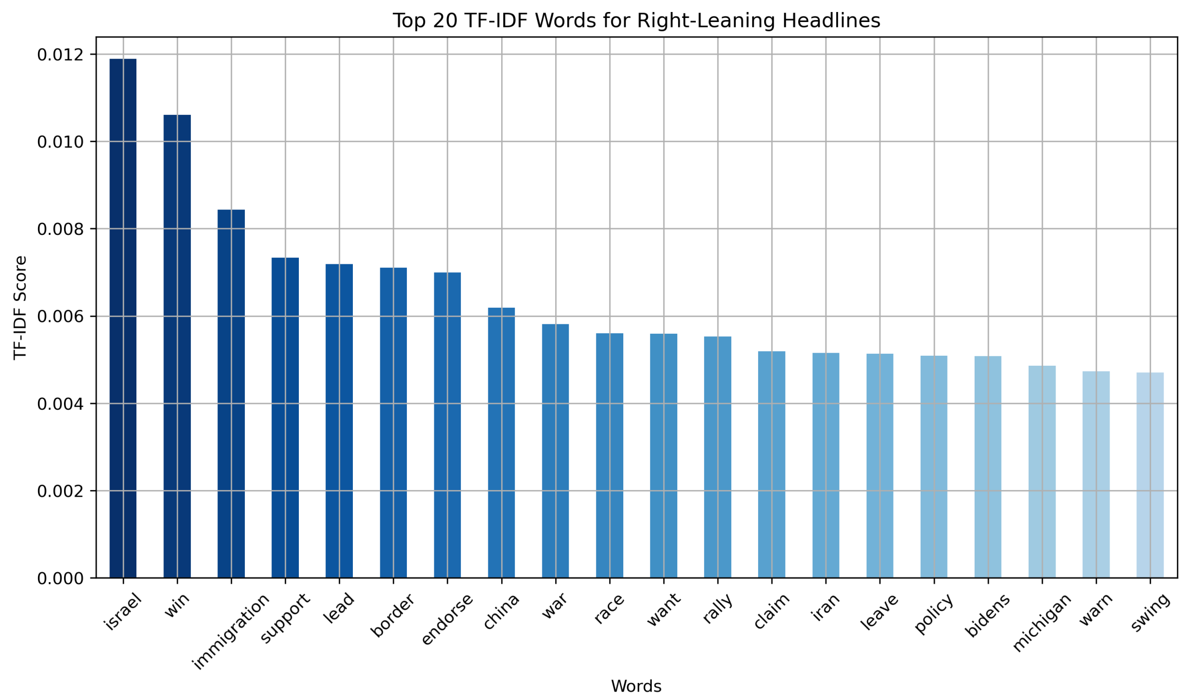

Figure 7: Tf-idf ranking of right-leaning headlines

4.2.1 Israel - Palestine topic

With the common topic “israel”, bigram combined with sentiment analysis can be studied to identify how the two leaning presents an identical topic. Key terms like “israel”, “palestine”, “gaza”, “hamas”, “idf”, and “jerusalem” were chosen for finding bigrams including them and their products of occurences and sentiment are shown in Figures 8 and 9 below.

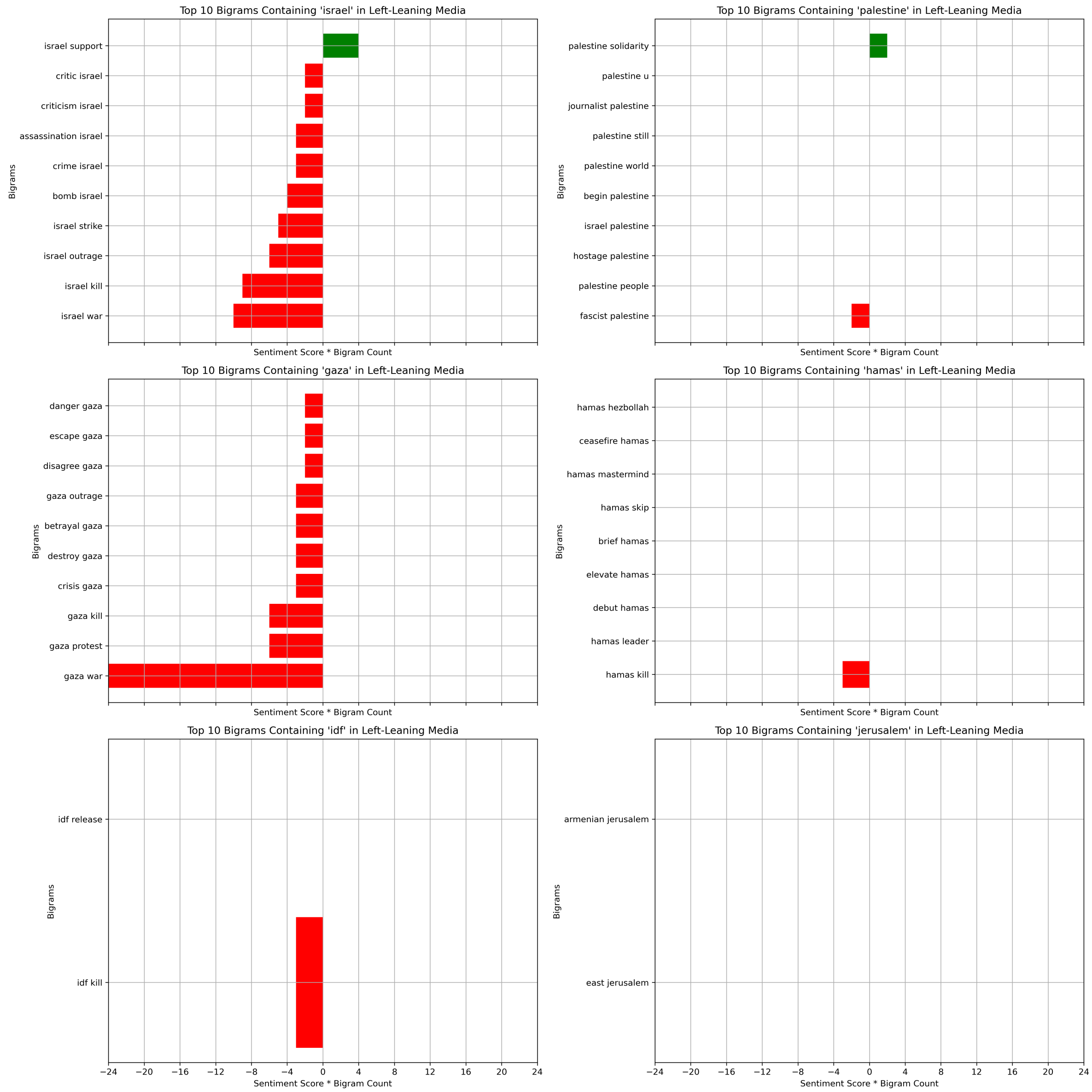

Figure 8: Bigrams occurence * sentiment of left-leaning headlines

From Figure 8, it can be observed that Israel, Gaza, and the IDF are used in more negative emotional terms, particularly in the context of violence (e.g., “kill”, “war”, “outrage”, “crisis”). In contrast, mentions of Palestine are more likely to include positive or neutral sentiment, indicating possible humanitarian or sympathetic framing. Hamas - a Palestinian militant organization, though associated with violence (e.g., “kill”), is not overly represented, possibly indicating cautious reporting by left-leaning outlets on this sensitive term. The scarcity of positive sentiment around Israel-related terms and the presence of emotional negativity (especially with Gaza and IDF) suggests framing aligned with criticism of Israeli military actions or pro-Palestinian advocacy.

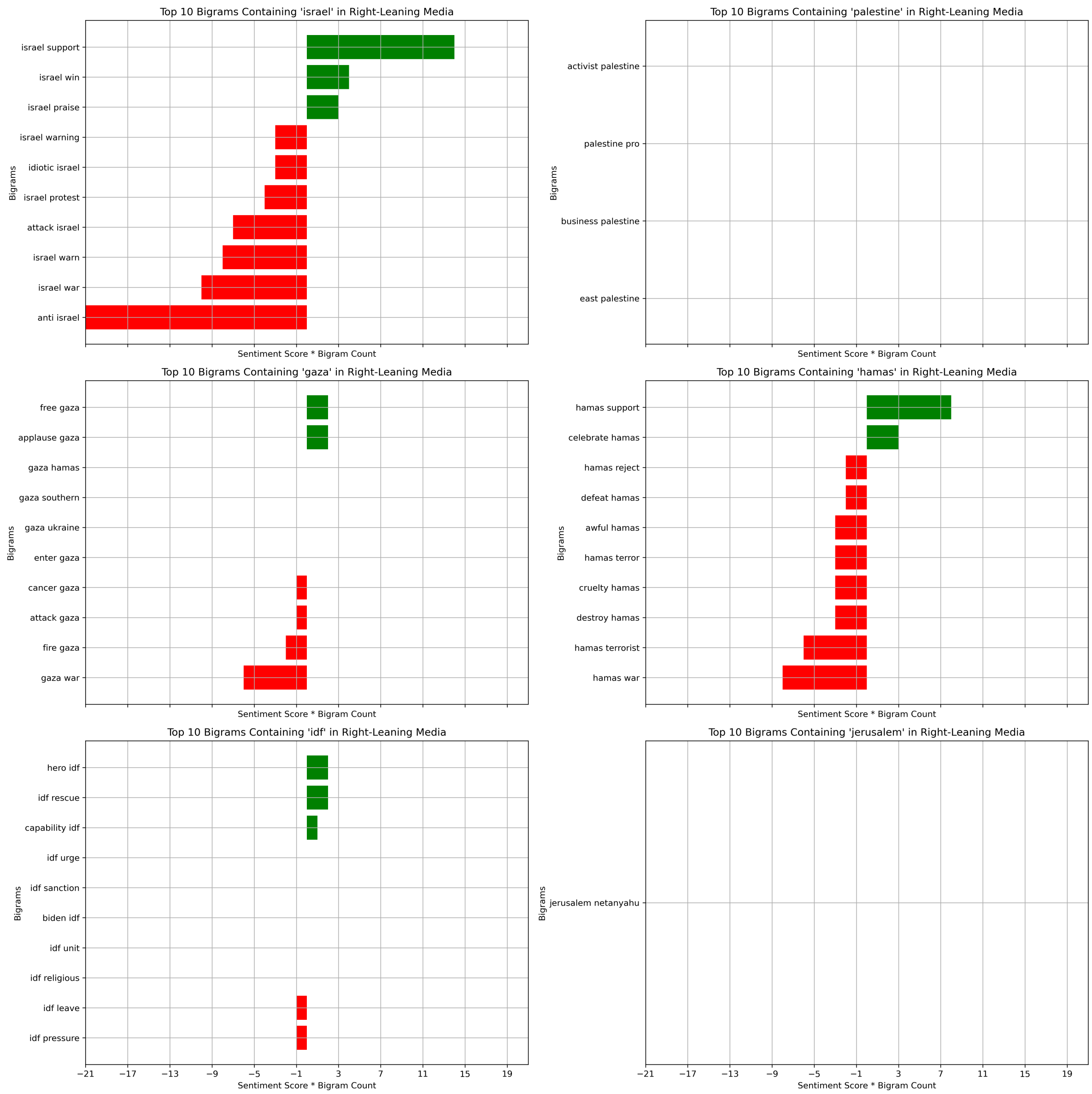

Figure 9: Bigrams occurence * sentiment of right-leaning headlines

On the other hand, Figure 9 shows that there are more positive framing on Israel and IDF, and overwhelming negative framing on Hamas. The presence of positive sentiment (e.g., “hamas support”) likely reflects lexicon ambiguity rather than genuine approval. Additionally, there was very limited emotional framing of Palestinian issues, and Gaza was presented more as a site of conflict, not primarily humanitarian crisis (“gaza hamas”, “gaza southern”, “enter gaza”). The constrast of viewpoints between left- and right-leaning outlets of this same topic showed how framing can be applied in political media, indicating that the counter-hypothesis for framing strategy could be rejected.

The bigram networks relating to the Israel-Palestine topic can also be displayed without the use of sentiment score, as shown in Figure 10 and 11. The networks reiterate the findings by showing that left-leaning media appear to focus more on the societal and moral dimensions of immigration. There is less clustering around militarized or enforcement language and more coverage that connects immigration to broader themes of justice, education, and support. (“lifelong school,” “future,” “safe strategy,” and “support outrage”). Topics associated with foreign policy spillover (e.g., “gaza,” “israel,” “hezbollah”) show up, indicating overlap between immigration and international conflict coverage. In contrast, the right-leaning network is more dense and centralized, especially around politically charged terms like: “israel,” “security,” “anti,” “terror,” “attack,” “bill,” “vote,” “protest,” and “administration.” Strong references to “gaza,” “hamas,” “hezbollah,” and “iran” co-occurring with immigration terms — suggesting a linking of immigration to national security and Middle East geopolitics. This also showed that immigration is framed as a threat or tie it to broader narratives of terrorism, foreign policy, and national security.

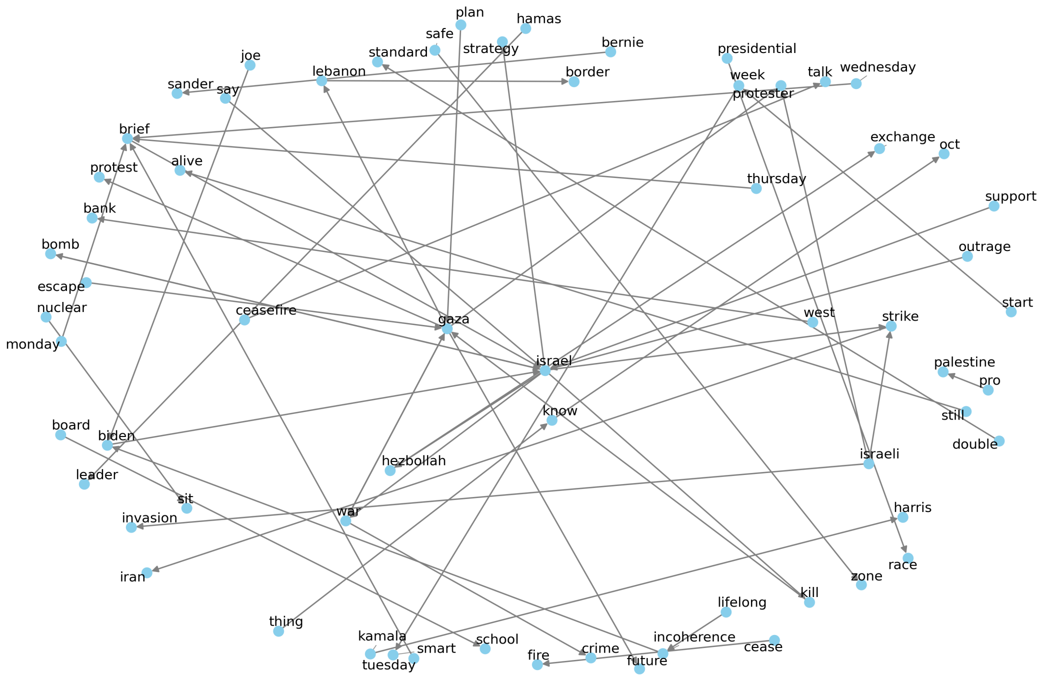

Figure 10: Bigram network of left-leaning outlets on Israel-Palestine situation

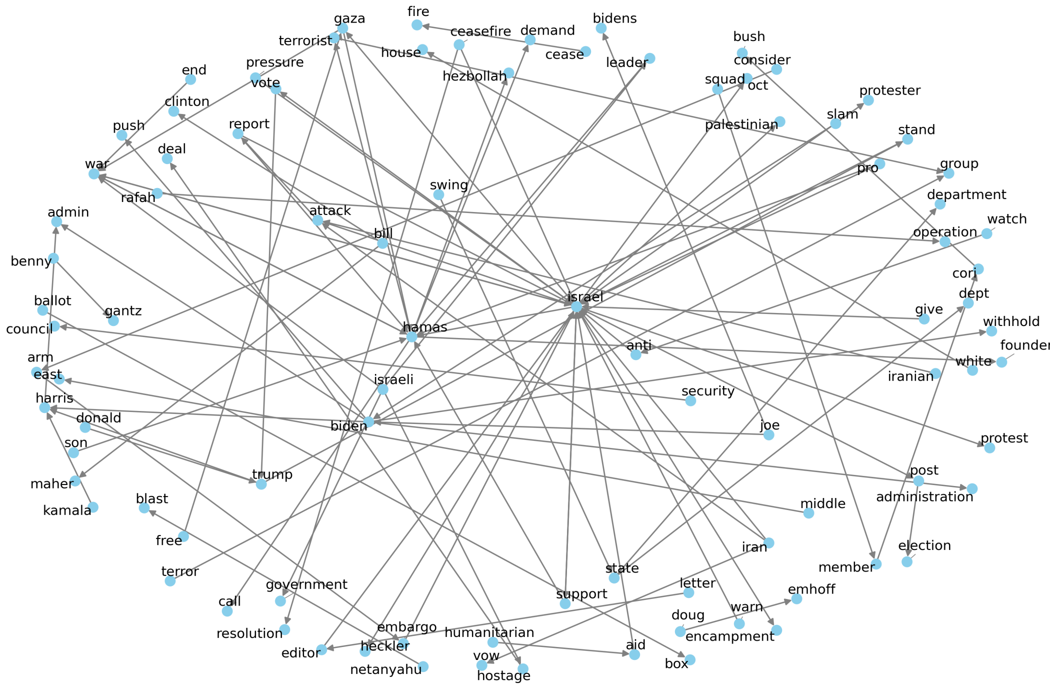

Figure 11: Bigram network of right-leaning outlets on Israel-Palestine situation

4.2.2 Immigration topic

Correlation technique can also be applied to identify different framing strategies of the two leanings. In Figure 12 below, words associated with "asylum", "migrant", and "refugee" strongly co-occur with terms like "seeker", "appointment", "solace", "displace", and various references to immigration logistics (e.g., "lottery", "app", "apply"). This indicates a focus on the bureaucratic, emotional, and life-saving dimensions of immigration. Terms such as "deportation" correlate with words like "graduate", "consequence", and "latinx", suggesting deportation is depicted as a threat to personal development and community stability. Meanwhile, "border" and "immigration" reflect some geopolitical references (e.g., "southern", "mexico", "lebanon", "crossing") but lack the intense threat-framing commonly seen in right-leaning discourse. Overall, these patterns support the hypothesis that left-leaning outlets frame immigration through lenses of compassion, opportunity, and systemic access.

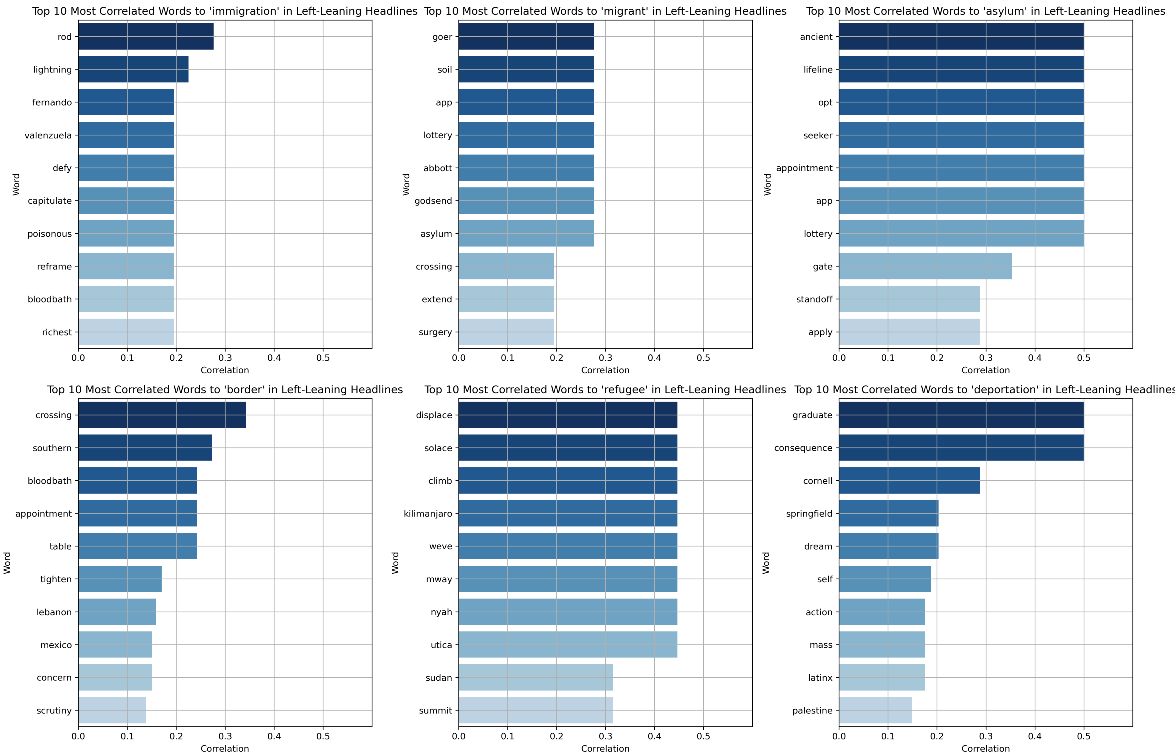

Figure 12: Correlated words of left-leaning headlines

Meanwhile, Figure 13 shows that the terms most associated with "immigration" include "illegal", "strain", and "repeal", indicating a narrative focused on burden and illegality. Similarly, words linked to "migrant" and "refugee"—such as "caravan", "cbp" (Customs and Border Protection), "crisis", "crime", and "massacre"—suggest associations with law enforcement, instability, or national security. The strong correlation of "border" with "czar", "patrol", and "crisis" further reinforces a conservative view of border control. The word "asylum" is notably tied to terms like "questionable", "falsely", and "minimize", which implies skepticism toward asylum claims. "Deportation" is associated with words like "mass", "priority", "halt", and "shock", all of which align with a more urgent framing of immigration removal policies. These co-occurrence patterns support the hypothesis that right-leaning media frames immigration through a lens of threat, control, and urgency, as opposed to humanitarian or rights-based narratives.

Networks for correlation with score larger than 0.5 can also be showed to display the constrat between the two leanings, as in Figures 14 and 15. In the left-leaning network, terms like "lifeline," "opt," "graduate," "appointment," "story," "apply," and "future" suggest a humanitarian or procedural framing of asylum and deportation. These associations often center on personal narratives, immigration policy logistics, and individual rights (e.g., "dream," "government," "controversial," "economic," and "application" processes). The presence of terms like “palestine,” “texas,” “troop,” and “threaten” signals occasional ties to geopolitical and domestic conflict contexts, but the dominant themes imply supportive or administrative language.

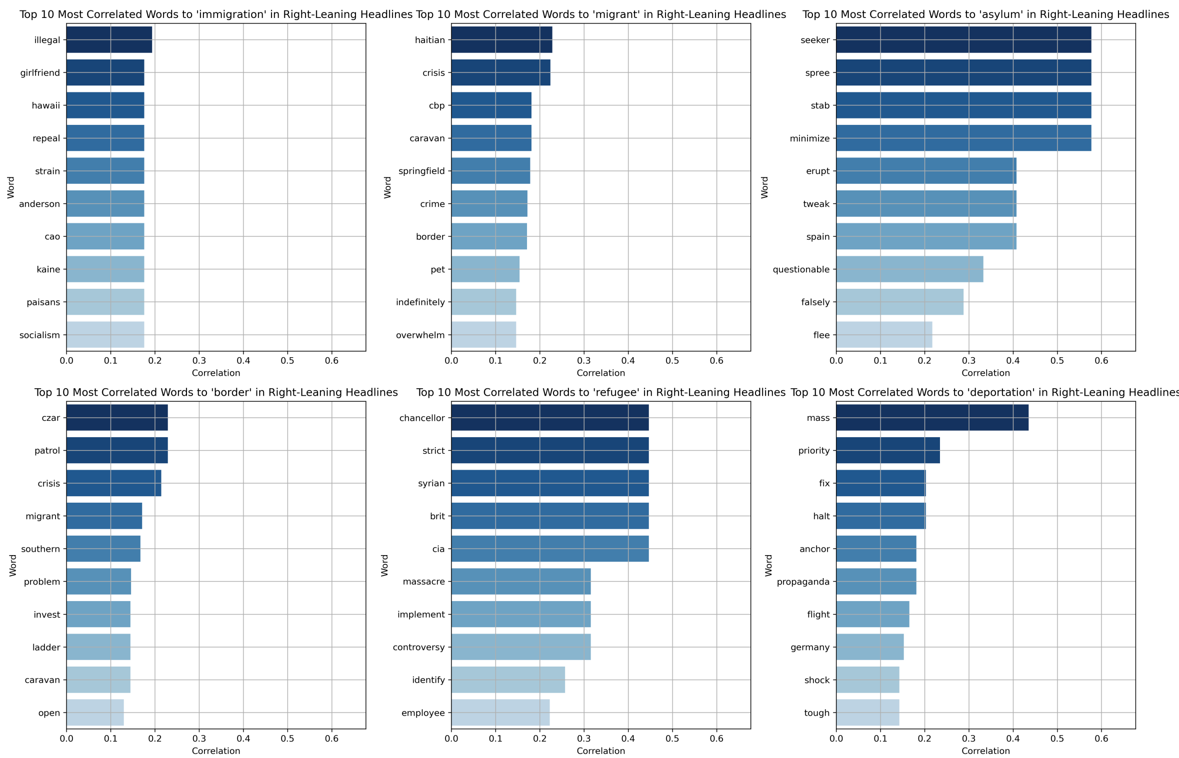

Figure 13: Correlated words of right-leaning headlines

The right-leaning network associated with “asylum” is considerably more charged and skeptical. Words like “stab,” “riot,” “spree,” “flee,” “minimize,” “questionable,” “falsely,” “rule,” and “coverage” suggest a narrative that links asylum-seeking with violence, dishonesty, or manipulation. The co-occurrence with terms like “cia,” “presidential,” “mass,” and “erupt” implies that asylum is also discussed in national security or political threat contexts. These patterns reinforce the earlier sentiment and correlation analyses.

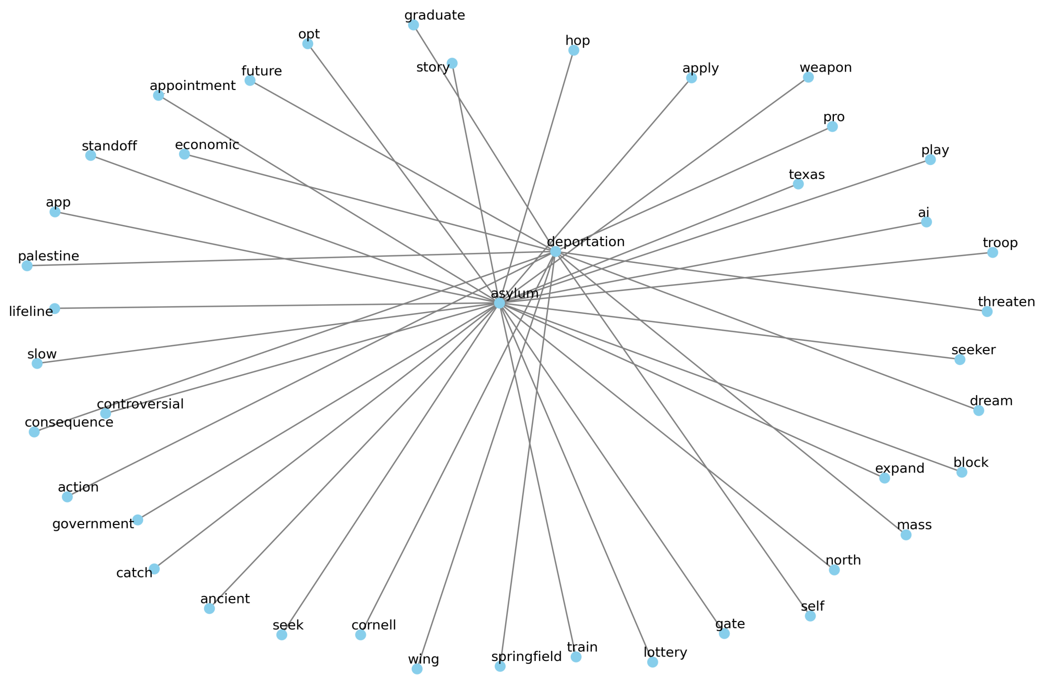

Figure 14: Correlation network of left-leaning with threshold 0.5

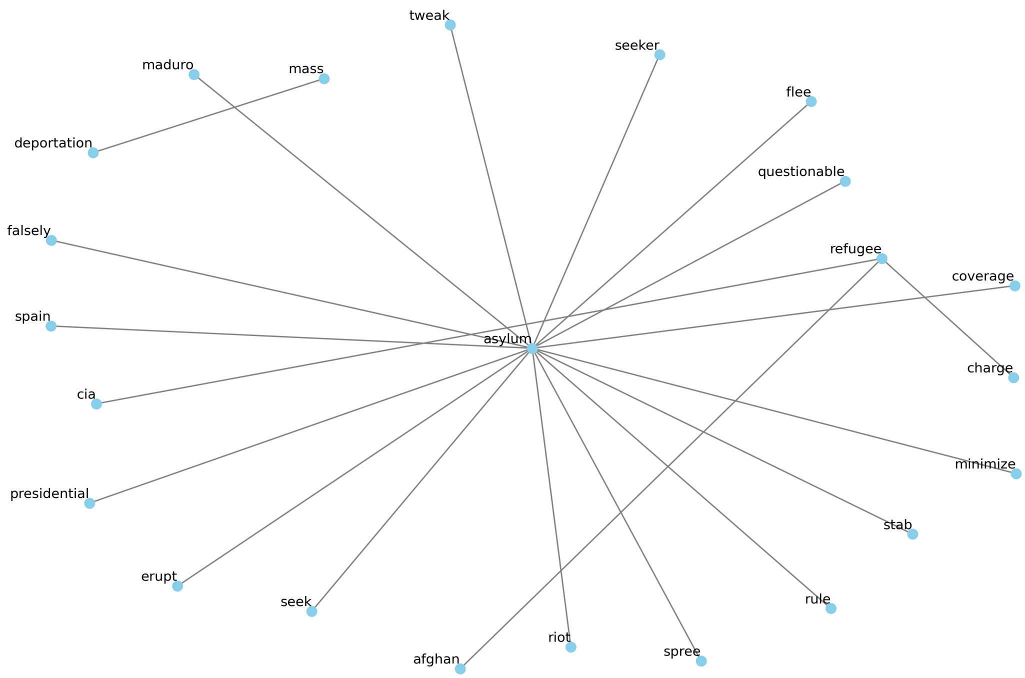

Figure 15: Correlation network of right-leaning with threshold 0.5

4.3 Topic modeling

By choosing the number of topic to be three, we obtained the following topics for the different leanings. Each subplot shows the top 30 most relevant terms for each topic, ranked by term frequency (blue bars) and TF-IDF score (green overlays), along with the percentage of tokens assigned to each topic across the corpus. These topics offer insight into the dominant themes emphasized in left-leaning coverage. For left-leaning outlets, the first topic (Topic 0), which accounts for 34.12% of the tokens, is clearly centered around the Israel–Palestine conflict and wartime coverage, shown in Figure 16. High-frequency terms such as "israel," "war," "gaza," "hamas," "attack," "air," and "strike" reflect extensive reporting on the military escalation in the Middle East. Topic 1 (32.90% of tokens) revolves around domestic social issues, particularly those related to protest, speech, and student activism. Words like "students," "speech," "campus," "protest," "antisemitism," "left," and "free" indicate strong media attention on U.S. college protests, cultural debates, and identity politics. The use of terms like "banned," "hate," and "free speech" points to ongoing tensions over expression and political ideology in educational institutions—an issue that gained visibility during the election season. Topic 2 (12.96%

of tokens) focuses on election-related discourse. Keywords such as “trump,” “biden,” “election,” “vote,” “campaign,” “rally,” “democrats,” and “republicans” highlight ongoing political rivalry and electoral developments.

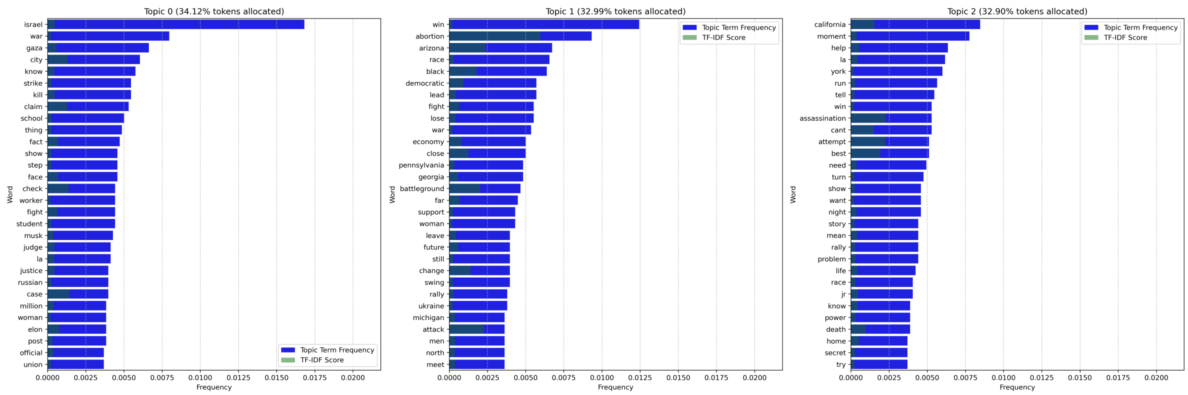

Figure 16: Topics of left-leaning headlines

Figure 17 showed the topic distribution for right-leaning outlets. Topic 0, which accounts for 34.13% of the total tokens, is dominated by keywords related to immigration and border control, such as “immigration,” “border”, “t-support”, “illegal”, “asylum,” and “migrant”. These terms reflect a framing focused on law enforcement, national security, and immigration reform—issues that are frequently emphasized in conservative political discourse. Topic 1 (34.77% of tokens) is primarily centered on the Israel–Hamas war and U.S. foreign policy. Frequent terms such as “israel”, “war,” “hamas”, “terror”, “attack,” “military,”, “antisemitism”, and “gaza” reveal heavy coverage of the conflict, but framed with a tone of security, terrorism, and support for Israel. Words like “rally”, “biden”, “antisemitism”, “support”, and “strong” highlight alignment with pro-Israel sentiment, which is consistent with long-standing Republican foreign policy positions. Topic 2 (31.10% of tokens) appears to focus on the domestic economy and cultural issues. It includes terms like “jobs”, “business”, “economy”, “inflation”, “biden”, “energy”, “vote”, and “crime”.

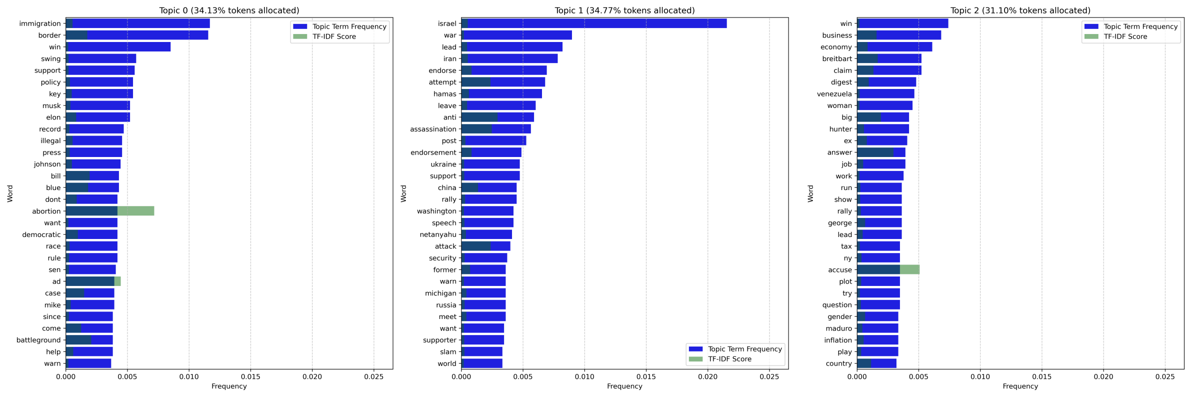

Figure 17: Topics of right-leaning headlines

The distribution of the topics overtime can also be showed in Figure 18 and 19 below. In the left-leaning media, the three leading topics are war and conflict (blue), social issues (green), and election and politics (red). These trends show a significant increase in coverage across all topics starting in July 2024, corresponding with the intense campaign season and the Democratic and Republican National Conventions. The war and conflict topic saw the most substantial rise, peaking in October 2024, likely due to crises such as the Israel–Hamas conflict. Social issues also rose steadily, reflecting the Democratic Party's platform. A sharp decline can be observed across all topics in November, which is expected, as the dataset ends in the middle of the month, shortly after Election Day (November 5, 2024).

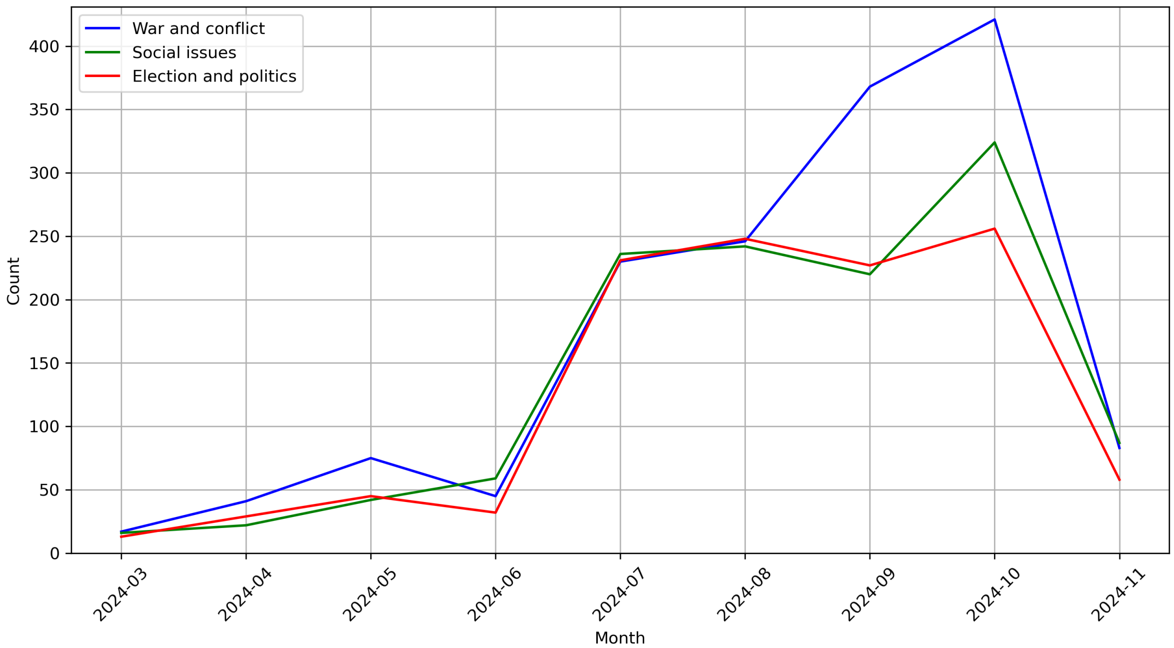

Figure 18: Topic trend of left-leaning headlines

In right-leaning media, the top topics were immigration and border security (purple), war and conflict (orange), and economy and cultural issues (brown). These topics were consistently prominent throughout the year, with a steep rise from July onward. Immigration and border security emerged as the most heavily covered topic, particularly in October 2024, aligning with campaign rhetoric focused on national security and migrant issues. The economy and cultural issues topic, maintained a stable presence and rose sharply as the election approached. These finding also supports the hypothesis of difference in topics covered by the two leaning, with left side focusing more on social issues and right side prioritizing immigration and economy topics.

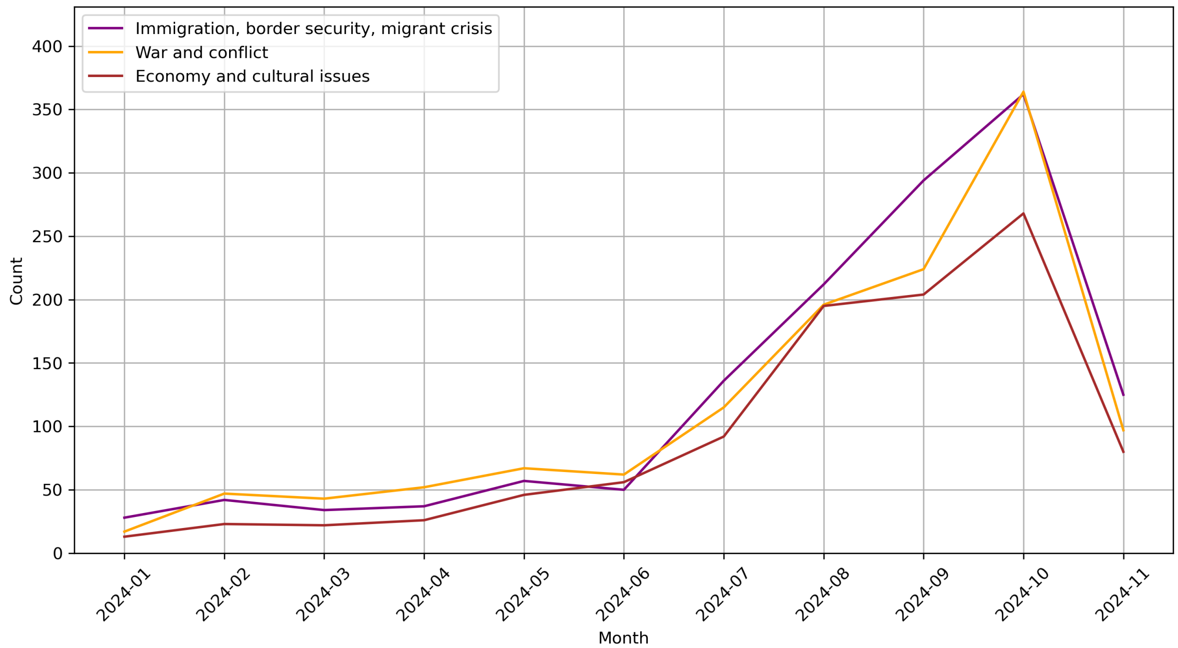

Figure 19: Topic trend of right-leaning headlines

5 Discussion and Conclusion

Based on the results above, this study provides substantial empirical evidence that media outlets with different political orientations exhibited distinct patterns in their coverage of the 2024 U.S. presidential election. Although both left- and right-leaning headlines contained mostly negative sentiment overall, the sentiment distribution over time revealed clear divergence. Additionally, the framing differences were further confirmed through bigram and co-occurrence analyses. Topic modeling via LDA further contrasts the two leanings: Left-leaning media primarily focused on war/conflict, domestic social issues, and election politics, while right-leaning media emphasized immigration, the Israel conflict, and economic/cultural issues. Thus, the overall counter-hypothesis that media across the political spectrum report on the election similarly can be rejected.

The findings of this study aligns with previous studies about framing strategies discussed in the introduction [14] [15]. Although the sentiments between the two leanings were quite similar, a focus of negative sentiment in right-leaning headlines new eletion days might indicate crowd manipulation discussed in [9] [10] [11] [12]. This study also serves as an example for analyzing different framing strategies using several techniques. Future study can apply correlation and bigram analyses on other topics, such as economy, climate, and gender issues. While having around 3000 headlines for each leaning, the headlines can contain noise or use sarcasm or dog whistles. More studies can focus more on quality of the data by or investigate different word choices in the headline. Another direction for future study would be to examine how coverage in strongly biased outlets compares to that of more centrist or neutral sources, particularly in terms of sentiment balance, topic diversity, and framing strategies. This could help assess whether centrist media serve as a moderating force in a polarized media environment or if they also exhibit subtle ideological tendencies in politically charged contexts.

A Appendix

https://github.com/minhNgnn/US-election-2024-headline-analysis

[1] Pew Research Center, "News platform fact sheet." https://www.pewresearch.org/journalism/fact-sheet/news-platform-fact-sheet/, 2024.

[2] Harvard University, "US media index – future of media project." https://projects.iq.harvard.edu/futureofmedia/us-media-index, 2024.

[3] A. Press, "Biden drops out of the 2024 race after debate stumbles, endorses harris as successor," *AP News*, 2024, Available: https://apnews.com/article/biden-drops-out-2024-election-ddffde72838370032bdcff946cfc2ce6

[4] K. Shepherd and G. Panetta, "Biden is dropping out of the 2024 presidential race. He is endorsing kamala harris." *The 19th News*, 2024, Available: https://19thnews.org/2024/07/biden-withdraws-2024-president/

[5] T. U. S. S. Staff, "Who won the US election? Results for president 2024 revealed," *The U.S. Sun*, 2024, Available: https://www.the-sun.com/news/12801091/who-won-the-us-election-president/

[6] Wikipedia contributors, "Timeline of the 2024 united states presidential election." https://en.wikipedia.org/wiki/Timeline_of_the_2024_United_States_presidential_election, 2024.

[7] M. E. McCombs and D. L. Shaw, "The agenda-setting function of mass media," *Public opinion quarterly*, vol. 36, no. 2, pp. 176–187, 1972.

[8] R. M. Entman, "Framing: Toward clarification of a fractured paradigm," *Journal of Communication*, vol. 43, no. 4, pp. 51–58, 1993.

[9] T. Groseclose and J. Milyo, "A measure of media bias," *The Quarterly Journal of Economics*, vol. 120, no. 4, pp. 1191–1237, 2005.

[10] J. M. Berry and S. Sobieraj, *The outrage industry: Political opinion media and the new incivility*. Oxford University Press, 2013.

[11] S. N. Soroka, "Negativity in democratic politics: Causes and consequences," *Cambridge University Press*, 2014.

[12] C. Budak, S. Goel, and J. M. Rao, "Fair and balanced? Quantifying media bias through crowdsourced content analysis," in *Proceedings of the 2016 ACM conference on computer supported cooperative work and social computing*, ACM, 2016, pp. 258–269.

[13] S. Iyengar, *Is anyone responsible? How television frames political issues*. University of Chicago Press, 1991.

[14] M. A. Baum and T. Groeling, "New media and the polarization of american political discourse," *Political Communication*, vol. 25, no. 4, pp. 345–365, 2008.

[15] N. J. Stroud, "Media use and political predispositions: Revisiting the concept of selective exposure," *Political Behavior*, vol. 30, no. 3, pp. 341–366, 2008.

[16] AllSides, "Media bias ratings." 2024. Available: https://www.allsides.com/media-bias/media-bias-ratings

[17] Ad Fontes Media, "Interactive media bias chart." 2024. Available: https://adfontesmedia.com/interactive-media-bias-chart/

[18] P. R. Center, "U.s. Media polarization and the 2020 election: A nation divided," Pew Research Center, 2020. Available: https://www.pewresearch.org/journalism/2020/01/24/u-s-media-polarization-and-the-2020-election-a-nation-divided/

[19] S. Bird, E. Klein, and E. Loper, "Natural language toolkit," in *Proceedings of the ACL 2009 conference: Short papers*, Association for Computational Linguistics, 2009, pp. 61–64.

[20] G. A. Miller, "WordNet: A lexical database for english," in *Proceedings of the 14th conference on computational linguistics*, Association for Computational Linguistics, 1995, pp. 39–45.

[21] S. M. Mohammad and P. D. Turney, *Crowdsourcing a word–emotion association lexicon*, vol. 29. 2013, pp. 436–465.

[22] F. Årup Nielsen, "AFINN: A new sentiment analysis baseline." https://github.com/fnielsen/afinn, 2011.

[23] G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval," *Information Processing & Management*, vol. 24, no. 5, pp. 513–523, 1988.

[24] K. W. Church and P. Hanks, "Word association norms, mutual information, and lexicography," in *Proceedings of the 27th annual meeting on association for computational linguistics*, Association for Computational Linguistics, 1990, pp. 76–83.

[25] D. Jurafsky and J. H. Martin, *Speech and language processing*. Prentice Hall, 2000.

[26] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of Machine Learning Research*, vol. 3, pp. 993–1022, 2003.