Journal of Computational Social Science (2022) 5:427–448
https://doi.org/10.1007/s42001-021-00130-y

RESEARCH ARTICLE

Using word embeddings to probe sentiment associations of politically loaded terms in news and opinion articles from news media outlets

David Rozado[1] · **Musa al-Gharbi**[2]

Received: 20 November 2020 / Accepted: 7 June 2021 / Published online: 14 July 2021

Abstract

This work describes an analysis of political associations in 27 million diachronic (1975–2019) news and opinion articles from 47 news media outlets popular in the United States. We use embedding models trained on individual outlets content to quantify outlet-specific latent associations between positive/negative sentiment words and terms loaded with political connotations such as those describing political orientation, party affiliation, names of influential politicians, and ideologically aligned public figures. We observe that both left- and right-leaning news media tend to associate positive sentiment words with terms used to refer to members of their own political in-group and negative sentiment words with terms used to denote members of their ideological outgroup. Outlets rated as centrist by humans display political associations that are often milder but similar in orientation to those of left-leaning news organizations. A weighted average of political associations by outlets' readership volume hints that political associations embedded in left of center news outlets might have larger societal reach. A chronological analysis of political associations through time suggests that political sentiment polarization is increasing in both left- and right-leaning news media contents. Our approach for measuring sentiment associations of words denoting political orientation in outlet-specific embedding models correlates substantially with external human ratings of outlet ideological bias ($r > 0.7$). Yet, specific sentiment associations are sometimes multifaceted and challenging to interpret. Overall, our work signals the potential of machine learning models derived from news media language usage to quantify the ideological bias embedded in news outlet content.

Keywords Word Embeddings · News Media · Political Bias · Media Bias · Polarization

✉ David Rozado
david.rozado@op.ac.nz

Extended author information available on the last page of the article

Introduction

A large amount of human culture is captured in textual artifacts. Consequently, content analysis of natural language corpora can provide rich insight into common associations prevalent within the social systems where the texts were produced [1, 2]. Corpora from news media articles represent an important source of information about the cultural landscape, as they contain salient and subtle semantic relationships, prevalent within the social context where said news articles and opinion pieces were written. Indeed, media coverage does not merely reflect socio-cultural values and perceptions, it also helps to shape them. In addition to conveying facts, coverage pushes viewers to understand the stories and various actors in particular ways through which stories are selected to receive attention or be ignored [3, 4], through metacommunicative frames [5] within and across texts, and by the attachment of emotionally and normatively laden words to various actors or actions [6].

Previous studies examining media organizations have found conflicting evidence regarding the existence, or lack thereof, of political bias in news media content [7]. Some studies have reported predominant Republican bias [8] in news media and other works have described predominant Democratic bias [9–11]. However, some of those findings have been disputed [12]. Several works have found bias in both left and right-leaning news organizations [13, 14] and evidence of growing political polarization in news media content [13]. Yet, others have only found favoritism towards political frontrunners [15], news outlet's endorsed candidates [16], or incumbents [17]. Some authors have also reported no evidence of partisan bias in news media [18–20].

Yet, when consumers of news media are asked about news outlets political bias, they do perceive that different news media organizations are politically biased [21]. However, human perceptions of political bias in news content can be highly subjective and often inconsistent. Previous work has found that when Republicans and Democrats watch the same news media coverage, Democrats tend to interpret the content as more favorable to the Republican candidate and vice versa [22]. Other work showed that presenting the same news media content but varying the label of its original source (CNN or Fox) changed humans' perceptions of political bias [23]. Thus, establishing the existence, or lack thereof, of bias in news media is not straightforward.

Methods for analyzing news media content using human coders, such as interpretative text analysis or qualitative coding, have been useful for the sociological analysis of news organizations. Unfortunately, such methods are constrained by their inability to scale to large corpora, and by low intercoder reliability when examining subtle themes. Computational content analysis techniques can circumvent some of the limitations imposed by content analysis using human raters.

Word embedding algorithms are a relatively recent language modeling technique that can capture latent semantic associations between words within a large volume of text into dense vector representations [24]. Embedding spaces have been shown to distil into their geometrical arrangements substantial amounts of factual information contained in the texts used to train the embeddings [25]. Some authors have reported that word embeddings also absorb many of the biases that are prevalent within a cultural system, as a result of embedding models being trained on culturally relevant language artifacts that are presumed to contain cultural biases, such as Wikipedia, Common Crawl, or repositories of contemporary books [26]. Others have used word embeddings trained on the Google Books Ngrams, a diachronic corpus of books encompassing a 100-year time period, to track the temporal evolution in embedding space of a reduced set of cultural associations and stereotypes around the dimensions of gender and ethnicity [27] as well as social class [28]. It has also been noted that the associations discovered by diachronic word embeddings correlate significantly with associations expressed by individuals both in historical and contemporary surveys [28]. Thus, a word embedding model trained on a specific corpus produced within a given cultural system can serve as a useful proxy to elucidate the idiosyncratic conscious and unconscious associations held or used by the group that produced the corpus of text.

In this work, we derive word embedding models from 27 million diachronic (1975–2019) news articles and opinion pieces published in 47 news media outlets popular in the United States. The list of outlets analyzed and their corresponding human ratings of political leanings was taken from the AllSides organization 2019 media bias chart v1.1 [29]. Temporal and outlet-specific embedding models were derived to probe for latent associations in each outlet and temporal interval between words with political connotations and sentiment lexicons. Political associations are measured by building outlet-specific political axes in embedding space that trace the spectrum between two poles representing distinct political orientation identities, such as conservatives-liberals or Republicans-Democrats [28]. We then project the vector representations of sentiment words onto those axes to measure positive/negative associations with the political categories represented in the poles of the axis. This method allowed us to characterize temporal and outlet-specific prevalence of positive or negative associations between political categories and sentiment valence.

When projecting words from popular sentiment lexicons onto a political axis, if the model displays a significant association of positive or negative words with either pole of the axis (each pole being representative of a political orientation), the result could be interpreted as outlet bias to maintain consistency with a common denotation of the term *bias* in the embedding models literature [26–28, 30]. However, as shown in the Results section, the associative space is often multifaceted and the causal roots of some associations are not always clear. Thus, we believe that the term *association* is sometimes more appropriate to describe linkages of uncertain origin in embedding space.

Using the aforementioned methodology, we first investigate contemporary and outlet-specific latent associations of positive and negative terms with the poles of axes representing distinct political affiliation. Next, we examine the relationship between said political sentiment associations embedded in news content and human ratings of outlets ideological orientation. We then use the chronological nature of news media articles to trace the dynamics of political associations in news outlets as they flux through time.

Methods

News articles content, temporal coverage, data preprocessing, and data availability

News and opinion articles from the outlets listed in Fig. 3 are available in the outlet's online domains and/or public cache repositories such as *Google cache*, *The Internet Wayback Machine* [31], and *Common Crawl* [32]. This work has not analyzed video or audio content of news media organizations, except when the outlet explicitly provides a transcript of such content in article form.

The temporal coverage of articles from different news outlets is not uniform. For most media organizations, news articles availability in their online domains or Internet cache backups becomes sparse as a function of articles' age. This is not the case for some news outlets, where availability of news articles goes back to the 1970s. The Supplementary Material (SM) illustrates the time ranges of article data analyzed based on news outlets articles online availability.

Textual content included in our analysis is circumscribed to the articles' headlines and main text and does not include other article elements such as figure captions. Targeted textual content was located in HTML raw data using outlet-specific XPath expressions. Tokens were lowercased prior to estimating embedding models. Markup language tags, URLs, nonalphanumeric characters, punctuation, digits, 330 common stop words and multiple spaces were removed prior to estimating word embeddings models.

All the analysis scripts and the diachronic word embedding models built from each of the 47 news media outlets analyzed in this work are available in the following repository https://doi.org/10.5281/zenodo.4797463.

For the purpose of reproducibility, we also provide in the above repository the articles' text used to train the news outlets embedding models with the caveat that outlets articles not accessible without a subscription have been excluded. Also, for the included articles, stop words have been removed and the remaining words have been randomly scrambled within a sliding window of size 10 to render the articles incomprehensible to a human reader. These steps have been taken to not infringe articles copyright. These preprocessing steps have only minor impact on Continuous Bag of Words (CBOW) word2vec and the results reported in this work are similar when using the scrambled articles text to train outlet-specific embedding models.

Deriving outlet-specific word2vec embedding models from news outlets articles text

We derived outlet-specific word embedding models at every 5-year time intervals within the 1975–2019 time range. The gensim [33] implementation of word2vec was used to train the embedding models. The continuous bag of words (CBOW) architecture performed slightly better than the Skip-Gram architecture in commonly used validation metrics, so it was used for all subsequent analysis.

For training the word embedding models, the following parameters were used: vector dimensions = 300, window size = 10, negative sampling = 10, down sampling frequent words = 0.0001, minimum frequency count of 5 (only terms that appear more than 5 times in the corpus were included into the word embedding model vocabulary), and number of training iterations (epochs) through the corpus = 5. The exponent used to shape the negative sampling distribution was the default 0.75.

Outlet-specific embedding models' performance across a range of commonly used semantic, syntactic, and analogy tasks was similar to popular pre-trained embedding models trained on corpora such as Twitter or Google books on similarity, association, and word analogy tasks [34]; see SM for detailed validation tests results.

Word projections on cultural axes and factual relationships in the empirical world

Semantically meaningful dimensions in vector space can be derived from word embedding models using vector algebra operations [28]. This technique exploits the inherent semantic structure of embedding spaces to extract cultural spectra such as ethnicity, gender, or socioeconomic status. For example, the vector difference $v_{\text{woman}} - v_{\text{man}}$, corresponding to the subtraction of the word vector representation for the term *man* from the word vector representation for the term *woman*, can be interpreted as a cultural/demographic axis (i.e., vector) in embedding space pointing from maleness towards femaleness.

Previous work has shown that the existence of culturally meaningful structural regularities in vector space can be used to solve analogical reasoning tasks such as *man* is to woman as king is to…? [24]. This is accomplished by carrying out the vector algebra operation $v_{\text{woman}} - v_{\text{man}} + v_{\text{king}} \approx v_{\text{queen}}$. That is, adding the gender vector pointing from maleness towards femaleness ($v_{\text{woman}} - v_{\text{man}}$) to the king vector results in a location in vector space very close to the solution of the analogy, the vector representation of the word queen. Thus, semantic meaning in word embedding models is not only contained in the direction of specific word vectors but also in the direction of derived vectors resulting from vector algebra operations between word vectors.

Other authors have shown that the cultural associations of a word in a corpus can be estimated by calculating the orthogonal projection of its word vector representation onto cultural dimensions of interest [28]. That is, the cosine of the angle between a word vector like v_{rugby} and a cultural axis vector such as gender v_G provides an estimate of the association in the corpus of the word *rugby* with the poles of the cultural axis: *men* and *women* (see SM for methodological details).

Several approaches have been proposed to measure the associations contained in the geometrical arrangement of words in embedding models, but they tend to generate similar results, since they are algebraically similar [26–28]. Here, following the methodology pioneered by others [28], we create axes denoting cultural/demographic categories such as gender, socioeconomic development, or political party affiliation in word embedding models. We then project onto those axes external data sets annotated with factual relationships about the empirical world and measure the correlation between terms of interest projections on cultural axes and their external factual data annotations. We do this to demonstrate that embedding models derived from individual news media outlets successfully absorb in their geometrical structure factual relationships about the empirical world.

We replicate previous results [26] about gender associations in popular pre-trained embedding models by constructing a gender axis in an embedding model trained on New York Times articles written between 2015 and 2019. This is done by averaging into a vector representation a set of male denoting words (i.e., man, men, male, and males) and subtracting the result from an aggregate vector of female denoting words (i.e., woman, women, female, and females) to create the gender axis. We then project terms describing professions onto the gender axis and correlate the landing locations of the profession terms projections onto the gender axis with the percentage of female representation in each profession. The results show that professions with a high percentage of female representation tend to project towards the female pole of the gender axis and male-dominated professions tend to project towards the masculine pole of the gender axis. Thus, mimicking the statistical relationship between professions and their percentage of female representation in the job market, see Fig. 1.

Similarly, we can project country names onto an economic development axis built by subtracting an aggregate of terms forming a poverty pole (poor, poverty, and underdeveloped) from a set of terms denoting a wealth pole (wealth, rich, wealthy, prosperous, and developed). Again, the projection values of country names onto the economic development axis is correlated with factual GDP metrics about each country. Using similar methodology to project car manufacturer brands onto a price axis mildly captures the average price of a car from the brand. Finally, projecting the names of current U.S. senators onto a political party affiliation axis captures the political party affiliation of the senators.

Projections of sentiment lexicons onto cultural axes with widespread accepted connotations of positivity and negativity

To validate that embedding spaces derived from news outlet content also capture many of the intuitive sentiment associations held by humans, four cultural axis were created in the same word embedding model derived from New York Times articles used in Fig. 1. The four cultural axes are Death-Life, Disease-Health, Dictatorship-Democracy, and malevolent to respectable historical figures. The set of words employed to build the poles of these axes are provided in the SM. To highlight patterns of association between the poles of the axes and sentiment words, we project onto these axes the Ideonomy positive/negative personality traits (IPT) sentiment lexicon that we use in the Results section.

Predictably, in the New York Times embedding model, positive words from the IPT sentiment lexicon tend to be associated with the poles denoting life, health, democracy, and widely admired historical figures (such as Martin Luther King, Gandhi, or Nelson Mandela). Negative words tend to be associated with the poles denoting death, disease, dictatorship, and malevolent historical figures

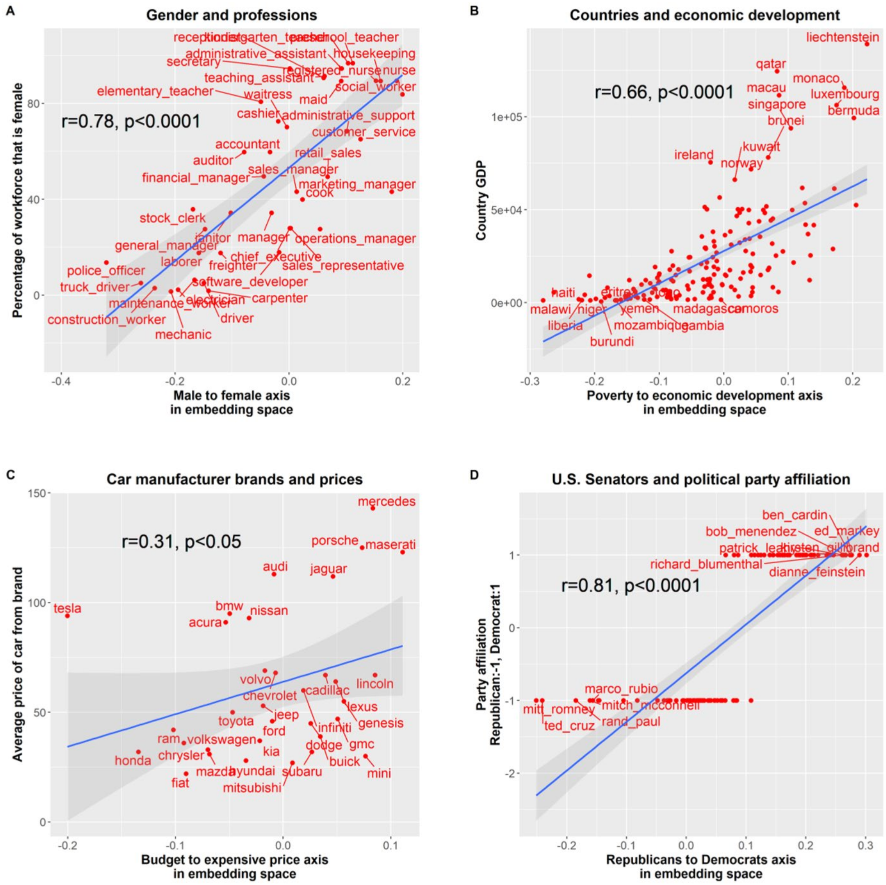


Fig. 1 A word embedding model trained on New York Times articles written between 2015 and 2019 captures in its geometric structure statistical metrics about factual empirical relationships such as the percentage of gender representation in different professions (**A**), countries' economic development (**B**), the average price of a car from a given brand (**C**), and the political affiliation of U.S. senators (**D**)

(such as Hitler or Stalin); see Fig. 2. The degree of correlation between IPT sentiment labels (positive or negative) and the corresponding words projection values on the studied cultural axes is moderate. Embedding models trained on news corpora other than the New York Times displayed similar patterns to those shown in Fig. 2. These results suggest that word embedding models trained on news outlet content capture many of the intuitive positive and negative sentiment associations held by humans. Thus, it is reasonable to use this technique to measure prevalent sentiment associations with respect to political orientation in embedding models derived from news outlets textual content.

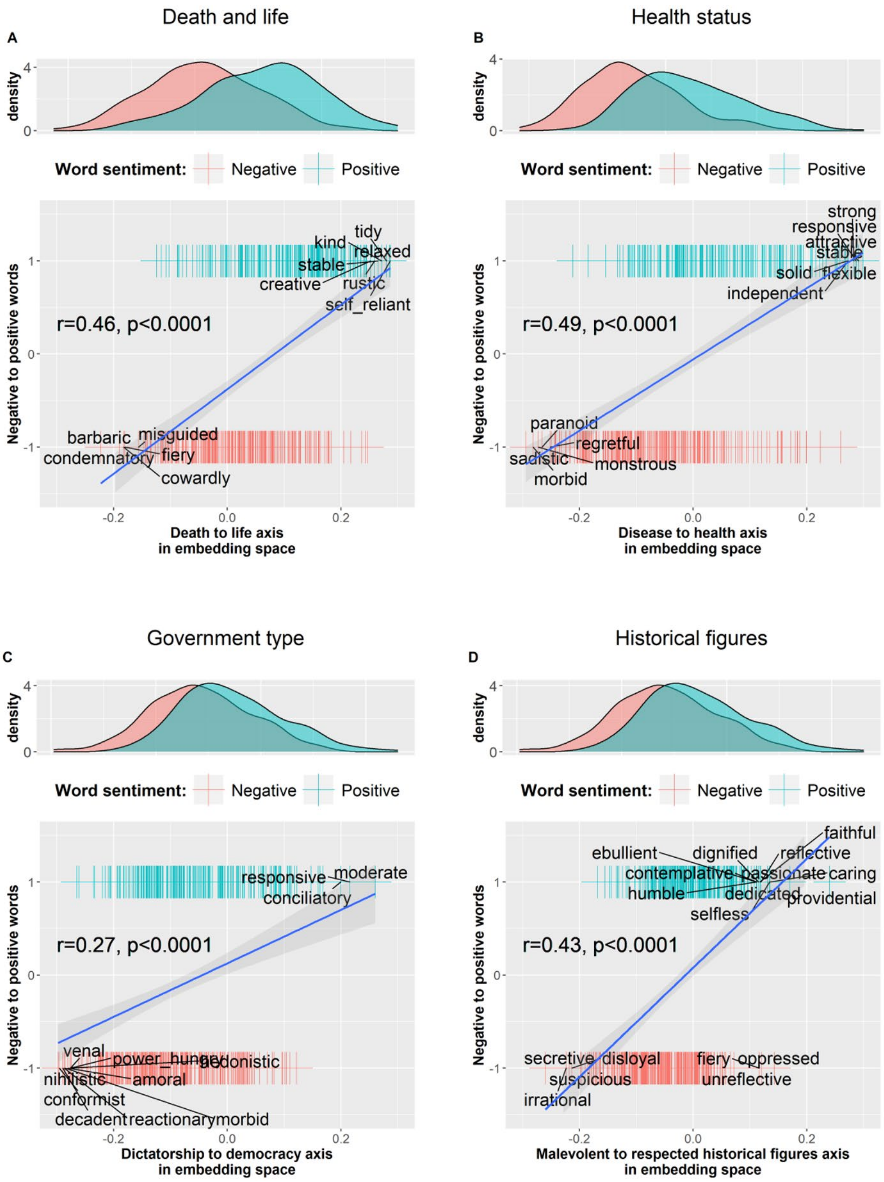


Fig. 2 The figure displays four cultural axes derived from a word embedding model trained with New York Times articles (2015–2019). The poles of all axes contain widely held positive and negative connotations: Death-Life (**A**), Disease-Health (**B**), Dictatorship-Democracy (**C**), or malevolent to respectable historical figures (**D**). The correlation between the projection values of IPT sentiment lexicon words onto these axes and their sentiment labels is shown as Pearson's r

Building cultural axes representative of political orientation

We built six axes representing political orientation for analysis purposes. A personal ideology cultural axis is built with terms such as *conservative*, *right-winger*, or *right-leaning* forming the conservative pole of the axis and terms such as *liberal*, *left-winger,* or *left-leaning* forming the liberal pole of the axis. The party affiliation cultural axis is built using terms such as *Republican*, *Republicans,* or *Republican Party* forming the pole representing Republican political affiliation. The Democratic pole is formed using terms such as *Democrat*, *Democrats,* or *Democratic Party*. The U.S. presidents' cultural axis is built with the names of all U.S. presidents since World War II, with each name being assigned to a pole of the axis according to presidents' political party affiliation.

The axis of ideologically oriented journalists is built using an external list from Politico [35] containing 35 influential journalists for each ideological orientation. Some of the journalists' names building the left-leaning journalistic pole are: Anderson Cooper, Rachel Maddow, Ezra Klein, Arianna Huffington, Nate Silver, Christiane Amanpour, Paul Krugman, Donna Brazile, Matthew Yglesias, Chris Matthews, Don Lemon, or Thomas Friedman. Some of the journalists building the right-leaning journalistic pole are Sean Hannity, Michelle Malkin, Bret Baier, Glenn Beck, Bill O'Reilly, Matt Drudge, Charles Krauthammer, Ann Coulter, Greg Gutfeld, or Tucker Carlson.

The U.S. senators' axis is built with the names of all Republican and Democratic senators in the U.S. Senate as of March 2020, with each member being assigned to a pole as a function of political party affiliation. Finally, the influential conservatives and liberals axis are built using two external lists from The Telegraph newspaper that ranked influential liberal [36] and conservative [37] public figures in the U.S. such as Supreme Court justices, high-ranking administration officials, politicians, and journalists. All names in this list that were already used in one of the previous axes were removed prior to construction of the axis to prevent redundant double counting. Details about the complete set of words forming the poles of each axis are provided as SM.

Results

Political associations in outlet-specific embedding models and external human ratings of outlet ideological bias

To measure political associations in news media outlets, we trained word2vec embedding models on contemporary outlet-specific corpora of news and opinion articles and quantified the degree of correlation between sentiment words labels (positive and negative) and the projection values of said sentiment words onto cultural axes tracing the spectrum of political orientation in the embedding models representative of each outlet. This methodology is similar to previous work measuring bias in word embedding models [28, 30].

Fig. 3 Over 10 million articles from 47 news media outlets within the 2015–2019 time range were used ▶ to create word2vec embedding models for each outlet. The barplot on top shows outlet-specific associations for 6 different political orientation axes. A positive value in the *y*-axis signifies associations favorable to left-wing denoting terms. A negative value signifies associations favorable to right-wing denoting terms. Average embedding associations in news outlets aggregated and color-coded by Allsides [29] human ratings of political orientation are represented with horizontal rectangles: blue (left-wing), sky blue (moderate left-wing), green (centrist), salmon (moderate right wing), and red (right wing). The average media bias reach into society weighted by monthly visitors to the top five news outlets web domains within each political orientation category, is shown with black horizontal dashed lines. The bottom scatter plots display the correlation between measurements of political associations in outlet-specific embedding models (*y*-axis) and AllSides human ratings of media bias (*x*-axis) across the six political axes analyzed

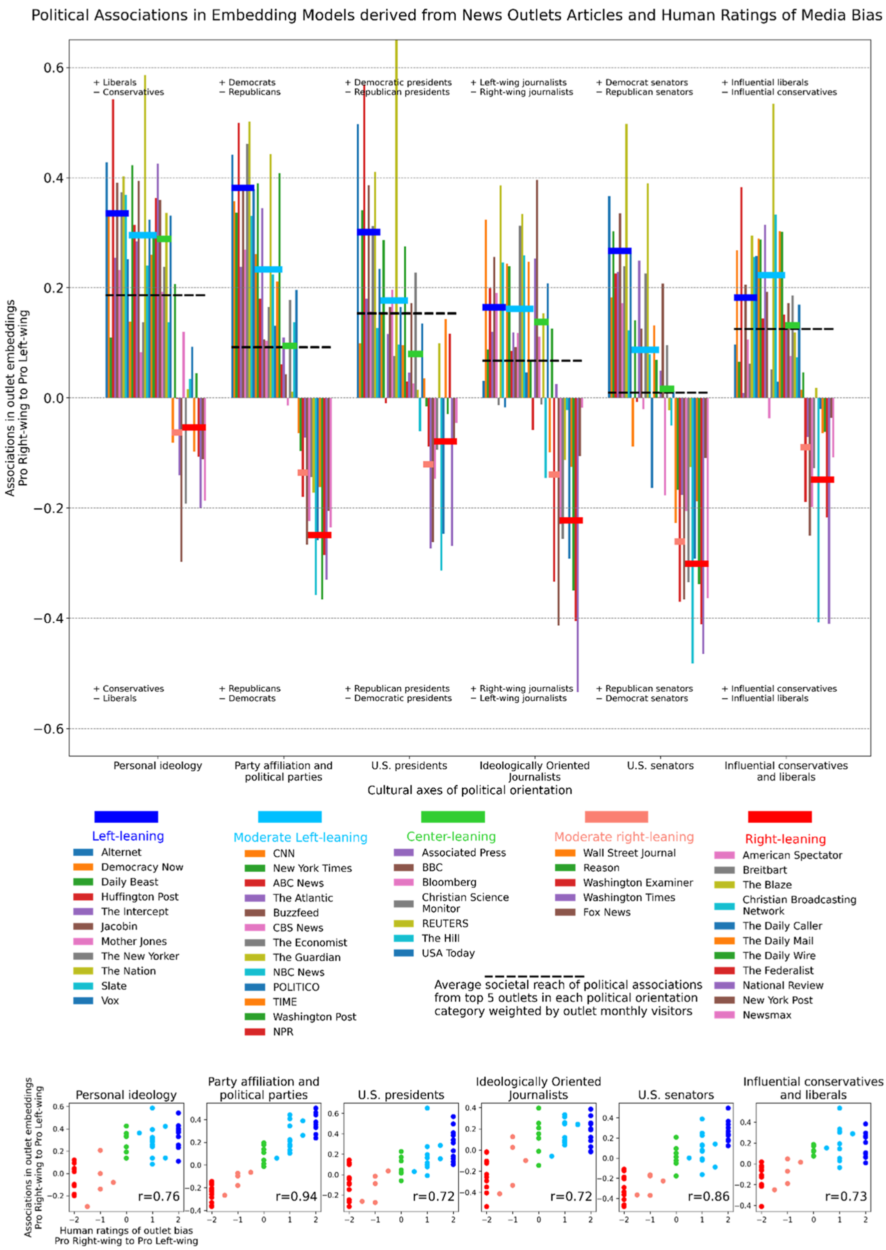
Overall, 19 different sentiment lexicons were tested to measure associations between positive/negative words and political orientation axes. The Ideonomy positive/negative personality traits lexicon ($N=526$) [38], or IPT for short, projection values displayed a slightly higher degree of correlation with human ratings of outlets ideological bias and it was used in the following analysis to measure political associations in embedding models derived from news media content. However, similar results were obtained when using any of the other 18 sentiment lexicon analyzed (see SM for detailed numerical results).

Systematic projection of the IPT sentiment lexicon onto six ideological axes for each outlet-specific embedding model trained on 2015 2019 articles and the resultant landing values correlation with the sentiment lexicon positive/negative labels revealed the association metrics shown in the *Y*-axes of Fig. 3.

Right-leaning news outlets tend to associate negative terms with liberals, Democrats, and left-wing public figures and positive terms with conservatives, Republicans, and right-wing public figures. An inverse association is observable in left-leaning news outlets. Outlets ranked as centrist by humans (see AllSides [29]) display associations that are milder but more similar in direction to those of left-leaning outlets than those of right-leaning media. Details about the bag of terms forming the poles of each axis are provided in the SM.

Weighting the associations of the top five most popular outlets in each political orientation category by monthly visitors to their online domains, according to web traffic metrics from SimilarWeb.com [39], we obtain a weighted average of written news media political associations reach into society (dashed black horizontal lines in Fig. 3). This metric appears to consistently show a larger reach of left-leaning media associations into society than right-leaning ones for most political axes. Yet, all two-sided single sample *t* tests for each political axis failed to reach statistical significance for the null hypothesis of neutral political associations' societal reach. However, using Fisher's method to aggregate the *p* values of each test resulted in a borderline combined Fisher statistic$=19.58$, $p=0.075$. An overall two-sided t test of all association measurements in all political axes for the null hypotheses that the weighted association reach into society is 0 resulted in $t(149)=3.1$, $p=0.002$.

Pearson's r correlations between political associations in outlet-specific embedding models and external human ratings of outlet ideological bias were substantial regardless of political axis analyzed (see scatter plots in the bottom of Fig. 3). This



hints at the ability of embedding models to consistently align political orientation axes in embedding space despite said axes being constructed with different sets of words whose only commonality is their factual political affiliation. The results described above are consistent regardless of sentiment lexicon or human ratings of media bias used, see SM for detailed numerical results. This is likely due to the large degree of sentiment annotations overlap between different lexicons and the high correlation between different human ratings of outlets political bias. Overall, these results suggest that measuring sentiment associations with politically loaded words in embedding models trained on outlet-specific articles' text can be used as a proxy to measure news outlets political bias.

A comprehensive visualization of the political orientation sentiment associations displayed by influential media outlets across 2 different political axes is shown in Fig. 4. Positive (in cyan) and negative (in salmon) words from 19 external sentiment lexicons (N = 15,704) are projected onto 2 political orientation axes of the seven most visited news media outlets in each broadly defined ideological orientation partition: left-leaning (top row), centrist (middle row), and right-leaning (bottom row). A clear tendency of left-leaning media outlets to associate liberals and Democrats with positive words (upper right quadrant of the plane) and conversely conservatives and Republicans with negative words (lower left quadrant of the plane) is apparent. A similar but milder trend is also apparent in some human rated centrist outlets. A few outlets display what appears to be mild or neutral associations. Several right-leaning media outlets display an opposite trend to left-leaning media outlets: a tendency to associate negative terms with liberals and Democrats and positive terms with conservatives and Republicans.

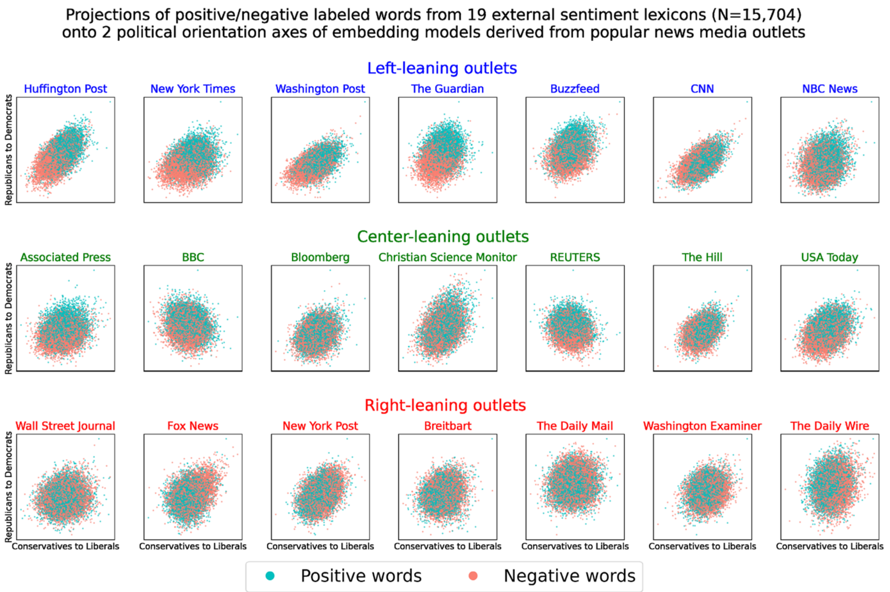


Fig. 4 Projections of positive and negative terms from 19 external sentiment lexicons (N = 15,704) onto 2 ideological orientation cultural axes derived from 21 popular news media outlets. Positive and negative sentiment words are color-coded in cyan and salmon, respectively. The upper right quadrant of each outlet plane represents liberals and Democrats in embedding space. The lower left quadrant represents conservatives and Republicans. AllSides labeled left-wing media occupy the top row, centrist media are located in the middle row, and right-wing media occupy the bottom row

Distinguishing the different facets of negativity

A common limitation of most sentiment lexicons is that the label *negativity* conflates a variety of different phenomena such as shame, anxiety, selfishness, suffering, sadness, fear, anger, aggressiveness, or dislike, under the umbrella category *negativity*. Thus, a news outlet that emphasizes the marginalization or suffering of a particular political group could display measurable negative associations with that group, as words that describe victimization and suffering are often labeled as '*negative'* in sentiment lexicons. Similarly, a news outlet that emphasizes the negative character traits of a political group by highlighting traits such as dishonesty, bigotry, violent behavior, intolerance or selfishness, will also display negative associations with that group, because such terms are also labeled as '*negative*.' Thus, it is important to disentangle what specific types of negative associations with a given political group are more salient in news outlet content.

To do that, we use the 3872 antonym pairs contained in the Wordnet lexicon and create cultural axes with each antonym pair. That is, we can use a Wordnet antonym pair, such as *unselfish-selfish*, to create a cultural axis that traces the spectrum going from altruism to selfishness in embedding space. We can then compare the cosine similarity between this axis and cultural axes describing political orientation. A high degree of similarity indicates that the word embedding model tends to more often associate each word in the antonym pair with a specific pole of the political axis.

One caveat about this methodology is that several antonym pairs in Wordnet do not have a clear positive/negative dichotomy such as for instance *centrifugal-centripetal*. To circumscribe the measurement to sentiment associations, we filter out words in antonym pairs that do not belong to a large and popular sentiment lexicon, the Harvard General Inquirer positive/negative lexicon ($N=3623$) or HGI for short [40]. That is, we keep only words from Wordnet antonym pairs that have been externally labeled as positive or negative. This leaves 357 antonym pairs with a clear sentiment valence.

The results of comparing the similarity between axes derived from the 357 Wordnet sentiment antonym pairs and a political orientation axis are shown in Fig. 5. It is apparent that most of the negative words that better align with the conservative pole of the political axis in left-leaning outlets denote negative character traits or behavior such as: *dishonest*, *hostile*, *unfit*, *thoughtless*, *malignant,* or *nasty*. Similarly, right-leaning outlets associate the liberal pole of the political axis with words such as *incompetent*, *insane*, *crooked*, *dangerous*, *evil*, *naïve,* or *wicked* that also do not characterize someone's suffering or victimization but instead emphasize negative character traits. It is not immediately clear to the authors the causal roots of such negative associations and whether other factors, beyond simple animosity against the ideological outgroup, could be at play in the emergence of these associations.

For example, a prominent left-leaning outlet in Fig. 5 strongly associates the word *illegal* with the conservative pole of the political orientation axis. It is uncertain whether this is due to that news outlet frequently associating in their written content conservatives with *illegal* actions or whether this association is due to the news outlet simply reporting or quoting conservatives frequent preoccupation and acute opposition to *illegal* immigration. This example illustrates that

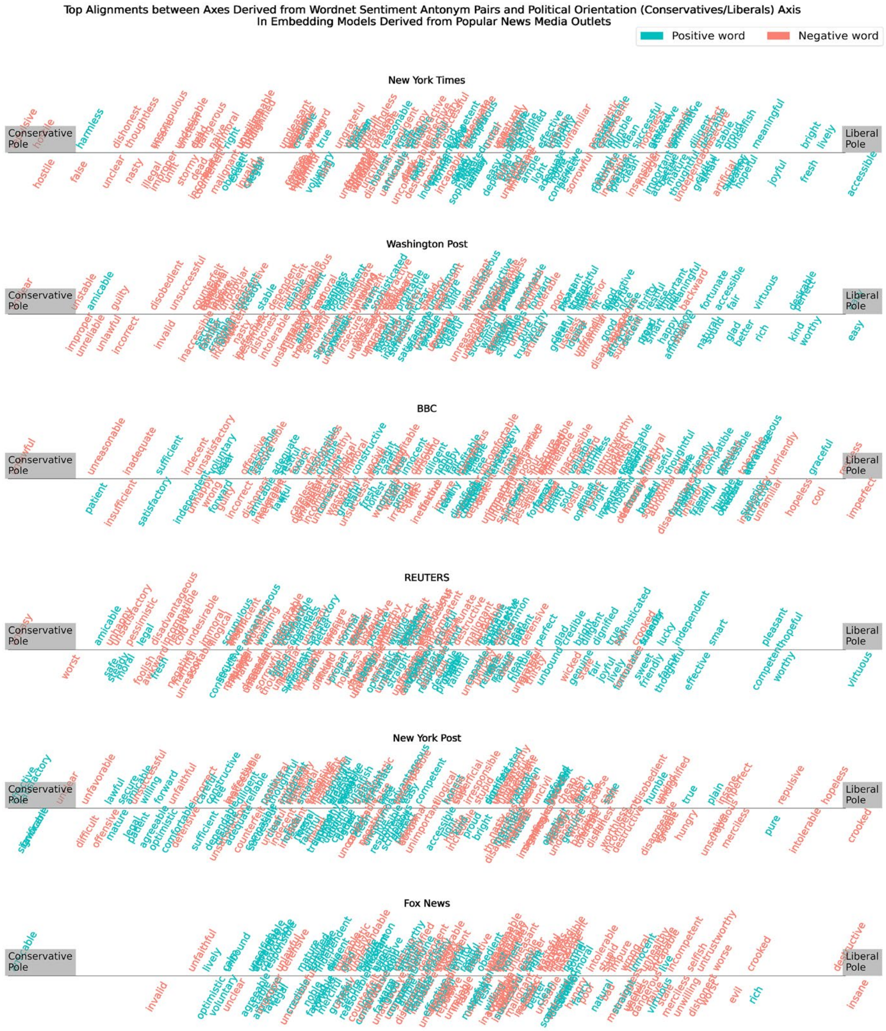


Fig. 5 Wordnet top 100 sentiment antonym pairs derived axes that better align in orientation with the conservatives to liberals cultural axis in embedding models derived from influential news media outlets. Words have been color-coded salmon and cyan to denote negative and positive labels. The words ‘progressive’ and ‘liberal’ have been left out from the figure, since they were outliers in several axes that distorted the visualization

the interpretation of specific semantic associations embedded in language models derived from news outlets corpora should proceed with caution. Although, on aggregate, our results suggest an overall strong correlation between political sentiment associations in news outlets corpora and outlets political bias, interested readers should resist the temptation to overinterpret specific individual associations as some of them are not always straightforward to decipher.

Comparing news media political associations between 2010–2014 and 2015–2019

The aforementioned results were measured in news and opinion articles from the 2015–2019 time range. To test whether outlets' associations are stable over time or instead reflect epiphenomena idiosyncratic to the socio-political moment between 2015 and 2019, we next compare political associations in media outlets for word embedding models trained on news outlets content from the 2015 to 2019 time interval versus embedding models trained on news outlets content from the 2010 to 2014 time range.

Figure 6 appears to display occasional evidence of growing polarization across political axes between 2010–2014 and 2015–2019. Paired *t* tests for each political axis (Bonferroni adjusted for multiple comparisons) show statistically significant growing sentiment polarization for terms denoting party affiliation and politically aligned influential public figures in left-leaning news outlets. Using Fisher's method to aggregate the *p* values of all paired *t* tests from left-leaning outlets resulted in a combined Fisher statistic = 51.30, $p = 8.25\text{e–}07$. Similarly, right-leaning news outlets appear to clearly have become more polarized when commenting on U.S. senators. Using Fisher's method to aggregate the p values of all paired t tests from right-leaning outlets resulted in a Fisher statistic = 23.63, $p = 0.023$. Overall, these results hint at an increase in polarization between 2010–2014 and 2015–2019. However, associations in 2010–2014 content display similar political leanings as those between

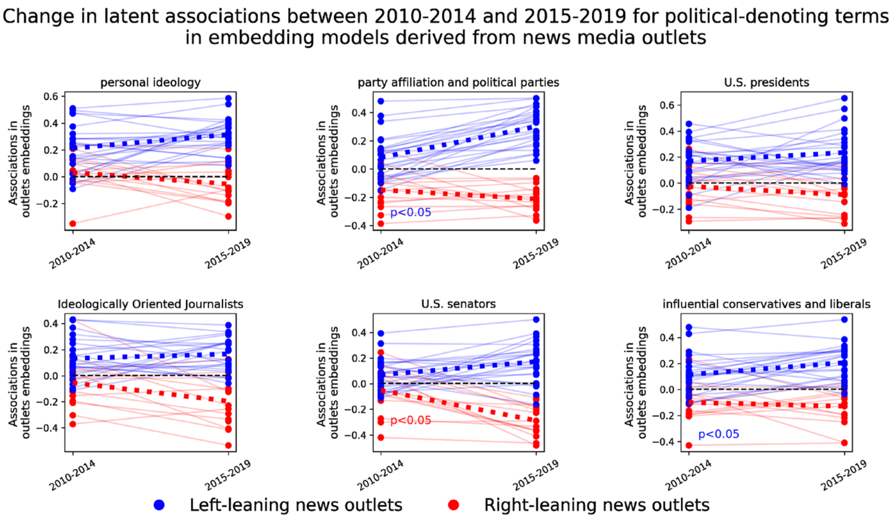


Fig. 6 Latent associations within news media outlets content in the time intervals 2010–2014 and 2015–2019 color-coded for human ratings of political orientation. Left-leaning and right-leaning media outlets are displayed as blue and red dots, respectively. Straight lines join bias measurements for the same outlet. Blue and red dotted lines represent the average bias for human rated left-leaning and right-leaning outlets. The dashed horizontal black line in the middle of each plot reveals the location of neutral association in embedding space. Significant p values for a paired *t* test (Bonferroni adjusted for multiple comparisons) between both time intervals associations in outlets embedding models are shown in the bottom left of each plot

2015 and 2019, as evidenced by consistently larger political association values in the *y*-axes for left-leaning outlets (denoting left-leaning preferential associations) than for right-leaning outlets. This suggests that the specific idiosyncrasies of the socio-political environment in the 2015–2019 time period have not been the only factor shaping the preferential political sentiment associations described in Figs. 3, 4 and 5 of this manuscript.

Long-term temporal dynamics of political associations in news media content

To study the dynamics of sentiment associations with politically loaded words for longer time frames, embedding models were generated for each outlet at 5-year time intervals as long as sufficient news articles content was available online. Thus, outlet-specific embedding models were generated for the following intervals: 1975–1979, 1980–1984, 1985–1989, 1990–1994, 1995–1999, 2000–2004, 2005–2009, 2010-2014, and 2015–2019. Representative online news articles availability for right-leaning outlets only starts in 1996. Therefore, trends for right-wing media outlets encompass a shorter timeframe.

Only three political axes formed with terms that have been fairly prevalent since the 1970s, such as *conservatives*, *liberals*, *Republicans,* or names of U.S. presidents, since World War II were analyzed. The three axes in Fig. 3 formed using contemporary politically aligned public figures (i.e., current US senators, influential journalists, and other public personalities) were not examined, since most of the names forming the poles of those axes were not prevalent in media discourse for previous decades.

Figure 7 displays sentiment associations' dynamics from 1975 to 2019, measured using outlets' content at 5-year intervals. The dashed horizontal black line in the

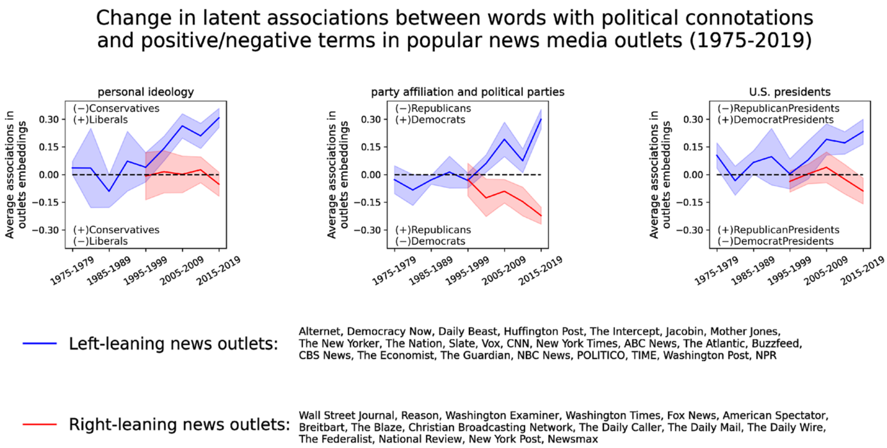


Fig. 7 Long-term analysis spanning several decades of latent sentiment associations in left-leaning and right-leaning news media outlets across several political axes. The dotted black line in each plot denotes a level of no preferential association of positive/negative words with either pole of the political axis. The shade of the trends shows the 95% confidence interval. The top left and bottom left of each plot indicate the positive/negative sentiment associations displayed by embedding models in that area of the chart

middle of each plot reveals the location of neutral association in embedding space. At this level of correlation between sentiment words labels and their projections on a political axis, the corpus displays no predilection to associate positive or negative terms with neither pole of a political axis. An association of negative sentiment terms with Pole 1 and/or positive sentiment terms with Pole 2 of an axis results in a positive correlation (above the dotted black line). This suggests bias against words that denote right-wing political sentiment. Conversely, an association of negative sentiment terms with Pole 2 and/or positive sentiment terms with Pole 1 of an axis results in a negative correlation (below the dotted red line). Such an association suggests bias against words that denote left-wing political sentiment.

Across the three ideological orientation axes analyzed, there appears to be a trend of increasing polarization in left-leaning outlets, see Fig. 7. That is, left-leaning outlets appear to have become more partisan over time and increasingly associate positive terms with words that denote their own ideological tribe and negative terms with words that denote their ideological outgroup. For each ideological axis in Fig. 7 from left to right, two-sided t tests with null hypotheses zero slope for left-leaning media were: $t(7)=4.62$, $p=0.002$, $t(7)=4.64$, $p=0.002$ and $t(7)=2.88$, $p=0.023$. For right-leaning news outlets, lack of articles data prior to 1995 is limiting for statistical robustness. Results for right of center news outlets are null for two of the ideological axes analyzed. There is however a hint of bias polarization in right-leaning news media due to increasing association of words denoting Republicans with positive words and words denoting Democrats with negative words. For each ideological axis in Fig. 7 from left to right, two-sided t tests with null hypotheses zero slope for right-leaning media were: $t(3)=-0.66$, $p=0.55$, $t(3)=-3.16$, $p=0.05$ and $t(3)=-0.97$, $p=0.40$.

In general, the results in Fig. 7 become more representative of the news media landscape as a function of time, since news media organizations and Internet cache repositories have less availability of news articles further into the past. That is, recent time intervals average a larger number of outlets. Hence, the trends in Fig. 7 could be due to the varying mix of outlets across time spans with perhaps more biased outlets more likely to have articles available only for recent years or more partisan outlets having emerged in recent years. Yet, mild evidence of polarization for left-leaning outlets is still apparent for a fixed set of 11 popular outlets with full article availability since at least the year 2000. For right-leaning outlets, however, the results are not significant and thus inconclusive, see SM for details.

Conclusion

Our results show that word embeddings algorithms can be used to computationally measure political sentiment associations in news media content and that such measurements correlate substantially with human perceptions of news media political bias ($r > 0.7$). That is, outlets that people perceive as left- or right-leaning tend to associate positive words with terms used to refer to members of their own political tribe and negative words with terms used to refer to members of their ideological outgroup. These results suggest that machine learning language models derived from large bodies of news media content can be successfully used to measure the ideological bias of news outlets.

Another relevant finding has been that some political associations in news media articles appear to display growing polarization with respect to how partisan media portray members of the 'other' faction. Partisan associations were already apparent in news content prior to 2015. However, over time, both left- and right-leaning outlets appear to be increasingly associating their political adversaries with negative terms and their political in-group with positive words in at least some of the axes analyzed. The trend however appears more consistent in left-leaning publications. The lack of online news articles content availability prior to 1996 for right-leaning outlets is limiting for deciphering right-wing-specific long-term trends.

Examining Figs. 3, 4, and 7 could lead to infer that political bias in left of center news media outlets is more acute than in right of center news outlets. Although this could conceivably be the case, it is not necessarily so. The inclusion of newswires from news agencies such as Associated Press or Reuters, that appear to display mild left-leaning associations (see Fig. 3), in both right and left-leaning media outlets, creates a confounding factor that could be ameliorating the pro right-leaning bias embedded in right of center news media outlets own content.

It is noteworthy that the increase in bias among left-leaning outlets seems to have begun in the mid-1990s—commensurate with the arrival of Fox News into the U.S. media landscape. Previous research has suggested that the emergence of Fox News polarized television audiences, not only solidifying conservatives within a right-wing media ecosystem, but also rendering the primary audiences of more traditional media outlets significantly more 'blue,' and eventually leading some of these outlets to adopt an explicitly left political posture in an appeal to cater to their new base [41].

Alternatively, the apparent asymmetric media polarization displayed in Fig. 7 could be a reflection of asymmetric political polarization. Some scholars have argued that although politicians and supporters of both parties have become more extreme in recent decades, these trends have been much more pronounced among Republicans than Democrats [42]. To the extent that the asymmetric political polarization narrative is true, the apparently growing bias in left-leaning media could be an artifact of attempting to discuss an opposition party whose political postures and positions left-leaning news media perceive as increasingly unreasonable.

A fundamental limitation of this work is that it is not clear how to precisely locate the neutral point of unbiased news. As highlighted previously by others [7], we lack a ground truth of neutrally covered real-world events that we could use as a reference to estimate unbiased news content. Hence, it is unclear whether some of the positive or negative sentiment associations shown in Fig. 3 reflect news media bias or a particular group using more negative language or producing a lot of detrimental news for themselves. That is, if a news outlet uses negative language to criticize a politician, it is uncertain whether the criticism reflects and accurate portrayal of the politician or a smear tactic. Regardless, such training data will push the embedding model towards negative associations with the targeted politician's name. Similarly, negative language describing victimization and negative character traits becomes indistinguishable to probe in embedding space, since both types of words are labeled as negative in standard sentiment lexicons. Recent NLP developments such as contextualized word embeddings [43] could help to mitigate some of these and similar technical limitations of static word embeddings.

Throughout this manuscript, we have described prevalent associations of positive/negative terms in news media content with words that denote political orientation. While the aggregate political association sentiment measures are validated by their high degree of correlation with human ratings of news outlets political bias, specific associations can sometimes be challenging to interpret. Beyond political animosity towards the ideological outgroup, other causal roots could be at play in creating such associations. Thus, we urge readers to exercise caution when attempting to interpret the complex entanglement of associations embedded in news media content due to the likely intricate and multifactorial roots of at least some of them. Finer-grained methodology will be needed to precisely characterize the causal roots of such linkages.

An important question arising from this work is what are the social consequences of news media bias and the increasing political polarization of news media content. On one hand, Americans trust in news media appears to be eroding [44], with this pattern being particularly acute among Republicans [45]. While most people would agree that fair and unbiased news reporting is an important element of a healthy civic culture, the commercial success of allegedly partisan news outlets suggests that there is a substantial market appetite for news content that supports the viewership's political predilections [46, 47]. The societal effects of increasingly partisan news media supply and consumption are therefore a relevant topic for future research.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s42001-021-00130-y.

Declarations

Conflict of interests The authors declare that they have no conflict of interest.

References

1. Neuendorf, K. A. (2001). *The Content Analysis Guidebook* (1st ed.). SAGE Publications Inc.
2. Rozado, D. (2019). Using word embeddings to analyze how universities conceptualize 'diversity' in their online institutional presence. *Society, 56*(3), 256–266. https://doi.org/10.1007/s12115-019-00362-9
3. *Setting the Agenda: Mass Media and Public Opinion*, 2 edition. Polity, 2014
4. Rozado, D. (2020). Prejudice and victimization themes in New York times discourse: a chronological analysis. *Academic Questions, 33*(1), 89–100. https://doi.org/10.1007/s12129-019-09857-7
5. Gorp, B. V. (2007). The constructionist approach to framing: bringing culture back in. *The Journal of Communication, 57*(1), 60–78. https://doi.org/10.1111/j.0021-9916.2007.00329.x
6. Gadarian, S. K. (2010). The politics of threat: how terrorism news shapes foreign policy attitudes. *The Journal of Political, 72*(2), 469–483. https://doi.org/10.1017/S0022381609990910
7. "Media Bias (Real and Perceived) and the Rise of Partisan Media," *Niskanen Center*, Nov. 06, 2017. https://www.niskanencenter.org/media-bias-real-perceived-rise-partisan-media/ (accessed May 03, 2021)

8. Grabe, M. E., Bucy, E. P. *Image Bite Politics: News and the Visual Framing of Elections*. Oxford University Press. Accessed: May 03, 2021. [Online]. https://www.oxford.universitypressscholarship.com/view/10.1093/acprof:oso/9780195372076.001.0001/acprof-9780195372076
9. Groeling, T. (2008). Who's the Fairest of them All? An empirical test for partisan bias on ABC, CBS, NBC, and Fox News. *Presidential Studies Quarterly, 38*(4), 631–657. https://doi.org/10.1111/j.1741-5705.2008.02668.x
10. Schiffer, A. J. (2006). Assessing partisan bias in political news: the case(s) of local senate election coverage. *Political Communication, 23*(1), 23–39. https://doi.org/10.1080/10584600500476981
11. Groseclose, T., & Milyo, J. (2005). A measure of media bias. *The Quarterly Journal of Economics, 120*(4), 1191–1237.
12. Gasper, J. (2011). Shifting ideologies? Re-examining media bias. *Quarterly Journal of Political Science, 6*, 357–370. https://doi.org/10.1561/100.00010006
13. Martin, G. J., & Yurukoglu, A. (2017). Bias in cable news: persuasion and polarization. *The American Economic Review, 107*(9), 2565–2599. https://doi.org/10.1257/aer.20160812
14. Larcinese, V., Puglisi, R., & Snyder, J. M., Jr. (2007). Partisan bias in economic news: evidence on the agenda-setting behavior of US newspapers. *National Bureau of Economic Research*. https://doi.org/10.3386/w13378
15. Waldman, P., & Devitt, J. (1998). Newspaper photographs and the 1996 presidential election: the question of bias. *Journalism & Mass Communication Quarterly, 75*(2), 302–311. https://doi.org/10.1177/107769909807500206
16. Barrett, A. W., & Barrington, L. W. (2005). Bias in newspaper photograph selection. *Political Research Quarterly, 58*(4), 609–618. https://doi.org/10.1177/106591290505800408
17. Kahn, K. F., & Kenney, P. (2002). The slant of the news: how editorial endorsements influence campaign coverage and citizens' views of candidates. *American Political Science Review., 96*(2), 381–394. https://doi.org/10.1017/S0003055402000230
18. *Media Bias in Presidential Election Coverage 1948–2008: Evaluation via Formal Measurement*. Accessed: May 03, 2021. [Online]. https://rowman.com/isbn/0739164740
19. Niven, D. (2003). Objective evidence on media bias: newspaper coverage of congressional party switchers. *Journalism & Mass Communication Quarterly, 80*(2), 311–326. https://doi.org/10.1177/107769900308000206
20. Dalal, S., Adlim, B., & Lesk, M. (2019). How to measure relative bias in media coverage? *Significance, 16*(5), 18–23. https://doi.org/10.1111/j.1740-9713.2019.01316.x
21. Mitchell, A., Gottfried, J., Kiley, J., Matsa, K. E. "Political Polarization & Media Habits," *Pew Research Center's Journalism Project*, Oct. 21, 2014. https://www.journalism.org/2014/10/21/political-polarization-media-habits/ (accessed May 06, 2021)
22. Eveland, W. P., & Shah, D. V. (2003). The impact of individual and interpersonal factors on perceived news media bias. *Political Psychology, 24*(1), 101–117. https://doi.org/10.1111/0162-895X.00318
23. Baum, M. A., & Gussin, P. (2008). In the eye of the beholder: how information shortcuts shape individual perceptions of bias in the media. *Quarterly Journal of Political Science, 3*(1), 1–31. https://doi.org/10.1561/100.00007010
24. Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., Dean, J. "Distributed Representations of Words and Phrases and their Compositionality," in *Advances in Neural Information Processing Systems 26*, C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2013, pp. 3111–3119. Accessed: May 11, 2017. [Online]. http://papers.nips.cc/paper/5021-distributed-representations-of-words-and-phrases-and-their-compositionality.pdf
25. Mikolov, T., Yih, W., Zweig, G. "Linguistic Regularities in Continuous Space Word Representations," in *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Atlanta, Georgia, 2013, pp. 746–751. Accessed: Mar. 19, 2019. [Online]. http://aclweb.org/anthology/N13-1090
26. Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science, 356*(6334), 183–186. https://doi.org/10.1126/science.aal4230
27. Garg, N., Schiebinger, L., Jurafsky, D., & Zou, J. (2018). Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences, 115*(16), E3635–E3644. https://doi.org/10.1073/pnas.1720347115
28. Kozlowski, A. C., Taddy, M., & Evans, J. A. (2019). The geometry of culture: analyzing the meanings of class through word embeddings. *American Sociological Review, 84*(5), 905–949. https://doi.org/10.1177/0003122419877135

29. "AllSides Media Bias Ratings," *AllSides*. https://www.allsides.com/blog/updated-allsides-media-bias-chart-version-11 (accessed May 10, 2020)
30. Rozado, D. (2020). Wide range screening of algorithmic bias in word embedding models using large sentiment lexicons reveals underreported bias types. *PLoS ONE, 15*(4), e0231189. https://doi.org/10.1371/journal.pone.0231189
31. Notess, G. R. "The Wayback Machine: The Web's Archive.," *Online*, vol. 26, no. 2, 2002, Accessed: Mar. 14, 2021. [Online]. https://www.elibrary.ru/item.asp?id=4370136
32. Mehmood, M. A., Shafiq, H. M., Waheed, A. "Understanding regional context of World Wide Web using common crawl corpus," in *2017 IEEE 13th Malaysia International Conference on Communications (MICC)*, 2017, pp. 164–169. https://doi.org/10.1109/MICC.2017.8311752
33. R. Řehůřek and P. Sojka, "Software Framework for Topic Modelling with Large Corpora," in *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, Valletta, Malta, May 2010, pp. 45–50.
34. Gladkova, A., Drozd, A., Matsuoka, S. "Analogy-based detection of morphological and semantic relations with word embeddings: what works and what doesn't.," in *Proceedings of the NAACL Student Research Workshop*, San Diego, California, Jun. 2016, pp. 8–15. Accessed: Mar. 25, 2019. [Online]. http://www.aclweb.org/anthology/N16-2002
35. Byers, D. "Twitter's most influential political journalists," *POLITICO*. https://www.politico.com/blogs/media/2015/04/twitters-most-influential-political-journalists-205510.html (accessed Aug. 02, 2020)
36. Harnden, T. "The most influential US liberals: 20–1," Jan. 15, 2010. Accessed: Aug. 02, 2020. [Online]. https://www.telegraph.co.uk/news/worldnews/northamerica/usa/6991000/The-most-influential-US-liberals-20-1.html
37. Harnden, T. "The most influential US conservatives: 20–1," Jan. 15, 2010. Accessed: Aug. 02, 2020. [Online]. https://www.telegraph.co.uk/news/worldnews/northamerica/usa/6990965/The-most-influential-US-conservatives-20-1.html
38. Gunkel, P. "Ideonomy Primary Personality Traits." http://ideonomy.mit.edu/essays/traits.html (accessed Aug. 08, 2020)
39. "Similarweb.com—Digital World Market Intelligence Platform," *SimilarWeb.com*. https://www.similarweb.com/ (accessed Nov. 19, 2020)
40. Stone, P. J., Dunphy, D. C., Smith, M. S. (1966). *The general inquirer: A computer approach to content analysis*. M.I.T. Press, 1966. [Online]. http://wjh.harvard.edu/~inquirer/
41. Enten, H. "How Roger Ailes Polarized TV News," *FiveThirtyEight*, May 19, 2017. https://fivethirtyeight.com/features/how-roger-ailes-polarized-tv-news/ (accessed May 20, 2021)
42. Mann, T. E. "Asymmetrical Polarization Undermined? Thoughts on the New Pew Research Center's Report on Political Polarization," *Brookings*, Nov. 30, 1AD. https://www.brookings.edu/blog/fixgov/2014/06/13/asymmetrical-polarization-undermined-thoughts-on-the-new-pew-research-centers-report-on-political-polarization/ (accessed May 20, 2021)
43. Devlin, J., Chang, M.-W., Lee, K., Toutanova, K. (2019). BERT: pre-training of deep bidirectional transformers for language understanding. https://doi.org/10.18653/v1/N19-1423
44. G. Inc, "Americans' Trust in Mass Media Sinks to New Low," *Gallup.com*, Sep. 14, 2016. https://news.gallup.com/poll/195542/americans-trust-mass-media-sinks-new-low.aspx (accessed May 06, 2021)
45. G. Inc, "Americans Remain Distrustful of Mass Media," *Gallup.com*, Sep. 30, 2020. https://news.gallup.com/poll/321116/americans-remain-distrustful-mass-media.aspx (accessed May 06, 2021)
46. al-Gharbi, M. "The New York Times' obsession with Trump, quantified," *Columbia Journalism Review*. https://www.cjr.org/covering_the_election/new-york-times-trump.php (accessed May 09, 2020)
47. al-Gharbi, M. "Cable news profits from its obsession with Trump. Viewers are the only victims.," *Columbia Journalism Review*. https://www.cjr.org/politics/cable-news-trump-obsession.php (accessed Mar. 22, 2021)

Authors and Affiliations

David Rozado[1] · **Musa al-Gharbi**[2]

[1] School of Engineering, Construction and Living Sciences, Otago Polytechnic, Dunedin, New Zealand

[2] Department of Sociology, Columbia University, New York, USA