2024

Tracking attention about COVID-19 vaccines on twitter and newspapers: A dynamic agenda-setting approach

Yi (Jasmine) Wang

Xjuli Wang

Jueman (Mandy) Zhang

Molu Shi

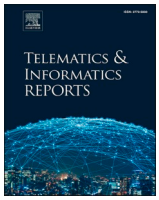
Wayne Wanta

Telematics and Informatics Reports 13 (2024) 100122

Contents lists available at ScienceDirect

Telematics and Informatics Reports

journal homepage: www.elsevier.com/locate/teler



Tracking attention about COVID-19 vaccines on twitter and newspapers: A dynamic agenda-setting approach

Yi (Jasmine) Wang [a,*], Xiuli Wang [b], Jueman (Mandy) Zhang [c], Molu Shi [d], Wayne Wanta [e]

[a] Department of Communication, University of Louisville, Louisville, KY 40292, USA
[b] School of New Media, Peking University, Beijing 100871, China
[c] Harrington School of Communication and Media, University of Rhode Island, Kingston, RI 02881
[d] College of Business, University of Louisville, Louisville, KY 40059, USA
[e] College of Journalism and Communications, University of Florida, Gainesville, FL 32611, USA

ARTICLE INFO

Keywords:
COVID-19
Vaccine
Agenda-setting
Topic modeling
LDA
Sentiment analysis
Granger causality
Twitter
Intermedia agenda-setting
Social media

ABSTRACT

This study delves into the intricate dynamics of agenda-setting between Twitter and elite news media concerning COVID-19 vaccines. A comprehensive dataset comprising 501,531 US-based, English-language tweets and 7,282 news headlines extracted from The New York Times and The Washington Post was collected from January 1, 2020, to April 30, 2021. To uncover the temporal evolution of content topics, Latent Dirichlet Allocation (LDA) was employed alongside sentiment analysis to gage corresponding valence levels. Granger causality tests were then conducted on the time series of topic sizes and valence scores from tweets and news headlines to explore the intermedia agenda-setting effects. The LDA analysis identified 13 topics, with Twitter discourse predominantly focusing on the top five ranked topics, while news headlines exhibited a more even distribution across all topics. The Granger causality tests revealed tweets-to-news Granger causality for four topics, news-to-tweets Granger causality for four topics, and mutual influence for the remaining five topics. Consequently, the directions of the agenda-setting effects varied depending on the specific discussions' topics. The findings indicated that elite news media wielded greater influence over socially impactful aspects of COVID-19 vaccination, while Twitter exhibited an agenda largely independent of elite news media, centering on highly personal facets of COVID-19 vaccination. Furthermore, the transfer of salience in topics was more pronounced compared to valence.

Introduction

The discussion surrounding the safety and effectiveness of vaccines, particularly those developed for COVID-19, has presented a significant challenge to public acceptance. These vaccines were created in an unusually short period of time, as highlighted by Ball [1], which has sparked widespread attention from the media and generated numerous discussions on social media. In this context, the role of news media has become increasingly crucial, as it serves as a vital channel for disseminating information related to health and environmental issues, exerting a substantial impact on the public's awareness and comprehension [2–4],. Hence, it is imperative to investigate how information regarding various aspects of COVID-19 vaccines was shared through diverse channels, such as reputable newspapers and popular social platforms, and to explore the reciprocal influence between these sources. Such analysis can provide valuable insights into effective communication strategies during public health crises.

While it has long been acknowledged that influential news media outlets such as The New York Times and The Washington Post have a significant impact on shaping public agendas and directing public attention [5,6], the emergence of participatory social media platforms, like Twitter, with their round-the-clock cycle, has disrupted the traditional one-way flow of agenda-setting power from elite news media to the general public [5]. A growing body of literature suggests that Twitter has become a vital and cost-effective resource for journalists in their quest for news and information, with Twitter discussions on specific topics even capable of triggering news coverage [7,8]. Consequently, the question of who sets the agenda in the digital age has been re-examined [9,10], yet the dynamics between elite news media and social media, particularly in the context of public health crises, have not received sufficient exploration.

This paper seeks to investigate the reciprocal relationship between

* Corresponding author.
E-mail address: jasmine.wang@louisville.edu (Y.(J. Wang).

https://doi.org/10.1016/j.teler.2024.100122
Received 29 June 2023; Received in revised form 13 November 2023; Accepted 24 January 2024
Available online 3 February 2024

the agenda set by elite news media, as demonstrated by their editorial choices in news selection, and the agenda shaped by public discourse on social media platforms, particularly Twitter discussions of COVID-19 vaccines, during the COVID-19 pandemic. Moving beyond the first level of agenda-setting, which primarily deals with the prominence of issues in the media, our research delves into the second level of agenda-setting. This involves examining the characteristics and sentiments surrounding an issue, suggesting that how these elements are presented in media discourse can also influence the public's perception and agenda [11]. Specifically, we aim to identify the areas where elite news media exert greater influence in determining the topics driving public discourse on COVID-19 vaccines and the instances where public discourse on social media, in turn, influences the reporting of elite news media regarding COVID-19 vaccines.

While traditional studies on agenda-setting have relied on manual coding techniques to identify key issues and attributes within news media coverage, recent advancements in natural language processing have introduced computational approaches, such as topic modeling and sentiment analysis, to analyze vast volumes of text and uncover latent information. Prior research has employed topic modeling to identify latent topics [12–15] and sentiment analysis [16] to assess the prevailing sentiments (positive, neutral, or negative) surrounding COVID-19 vaccine discourse on Twitter [12,13,17]. These descriptive studies have hinted at the influence of news media on Twitter discussions regarding COVID-19 vaccines [12,13,17]. However, no statistical analysis has been performed to substantiate this influence, and the impact of Twitter discourse on traditional news media remains unclear.

Employing advanced natural language processing (NLP) techniques, this research investigated the evolving patterns over time regarding the topics and emotional tone surrounding COVID-19 vaccines on Twitter, as well as within the coverage of two esteemed newspapers, namely The New York Times and The Washington Post, renowned for their influence in shaping the agendas of other media outlets [18]. Additionally, Granger analysis [19,20] was employed to explore the reciprocal relationship between the public discourse on Twitter and the agenda set by news media, with a focus on each specific topic under examination.

The current study aims to contribute to agenda-setting research by building upon the foundational work of earlier studies, extending the focus beyond the first level of agenda-setting, and utilizing automated coding compared to traditional manual methods. By integrating advanced natural language processing (NLP) techniques to analyze both the cognitive and affective dimensions of media content, this research offers a more sophisticated and nuanced analysis of the agenda-setting process. This methodological innovation facilitates a deeper understanding of how specific topics gain prominence and how they are emotionally framed, which is crucial in the context of public health crises where information and sentiment can rapidly change. Furthermore, applying Granger causality analysis to trace the directional influence between the elite media and Twitter, provides empirical evidence of the reciprocal nature of agenda-setting in the digital age. It also provides a framework for future research to explore the dynamic of agenda-setting effects in other contexts.

Literature review

The dynamic of agenda-setting

Agenda-setting refers to the influence exerted by mass media on the public's perception of issue salience [5]. Building upon this traditional framework, subsequent studies have explored factors that serve as precursors to the media agenda [21,22]. One crucial factor is intermedia agenda-setting, which usually pertains to the agenda-setting effects of various mass media on one another [6,18,23,24]. Scholars have argued that intermedia agenda-setting effects lead to content convergence and homogenization, as journalists tend to align their perspectives by validating the newsworthiness of topics through peer scrutiny, particularly among elite members [6,25–27].

However, the rapid emergence of social media platforms like Twitter challenges the traditional norms of professional journalism due to extensive audience participation and the expectation of a continuous news cycle. While early research on web logs indicated that social media might produce distinct agendas compared to traditional media [28], the agenda-setting effects between news media and social media can be reciprocal and mutually influential, contingent upon contexts and issues [29–33]. For example, Vargo and Guo [34] found that, in many cases, online partisan media are taking the lead in setting the media agenda, overshadowing traditional news sources like The New York Times and The Washington Post. Nevertheless, these traditional outlets still significantly shape the media narrative on specific topics, such as health care. This aligns with Vargo et al.'s [35] observation that traditional elite media exert a stronger agenda-setting influence on ongoing debates. Conversely, new media platforms have shown a notable impact in directing the media focus on topics like civil liberties, poverty, and religion, demonstrating their potential to amplify diverse perspectives [35]. Thus, further research is necessary to explore the varying impacts of different types of issues, such as those that are more or less obvious to the public [36] and those characterized by certainty versus uncertainty [37].

The agenda-setting effects manifest across multiple levels. The first level pertains to the transfer of issue salience [5]. The second level involves the transfer of attribute salience associated with specific issues [6,27,38–40]. These attributes can be categorized into cognitive and affective dimensions [38,39]. Cognitive attributes entail structured information about an issue, while affective attributes encompass emotional reactions oftentimes measured through valence [38,39]. For instance, the effectiveness of vaccines represents a cognitive attribute within the context of COVID-19 vaccines, while the positive, neutral, or negative sentiment associated with the content embodies an affective attribute. The third level of agenda-setting posits the transfer of salience regarding networked relationships among issues and attributes [41,42]. It suggests that by mentioning or linking different issues or attributes together, the audience's memory network can be activated, or new connections can be established. The choice of which level of agenda items to assess in agenda-setting research is both conceptual and methodological [43]. Su and Xiao [44] systematically reviewed the landscape of intermedia agenda-setting (IAS) studies from 1997 to 2019, revealing a predominant focus on the first level of agenda-setting. This level, which concerns the salience of issues within the media, was typically examined through content analyses that relied on manual coding techniques. Their comprehensive analysis underscores a critical call to action for future research to embrace a broader array of contexts and to leverage more sophisticated, innovative methodological approaches.

Our study focuses on the second level of agenda-setting, operationalizing cognitive attributes as topics related to COVID-19 vaccines and affective attributes as content valence. Furthermore, our research expands the examination of agenda-setting dynamics within the context of public health. This approach contributes to the theoretical advancement of agenda-setting literature and offers practical insights into the communication strategies that can effectively inform and engage the public on vital health-related issues.

Twitter as a source of public discourse during health crises

Twitter, as a prominent social media platform and a valuable source of text-based public discourse, has been extensively examined to gain insights into public reactions during public health crises. Previous research has explored Twitter's role in capturing public sentiments and responses during events such as the COVID-19 pandemic [45], the 2009 H1N1 pandemic [46], and the Ebola outbreak [47]. Public health researchers and stakeholders have utilized text-based public discourse on Twitter to identify risk factors [48] and monitor public perceptions of

health crises [46,47] as well as evaluate the effectiveness of health intervention measures [13,49].

In addition, Text-based public discourse on Twitter has also been examined from the dynamic agenda-setting perspective [9,10,39]. While most of these studies have primarily focused on political news agenda-setting, these inquiries have illuminated the intricate interplay between Twitter's agenda and that of traditional news media in notable events such as the 2019 Hong Kong anti-extradition bill movement [50] and the 2016 U.S. presidential election [39]. This perspective has also been applied to contexts like climate change [9] and natural disaster coverage [10]. However, there is a notable scarcity of research on the reciprocal influence between traditional media and Twitter in the realm of public health crises. A recent study by Shi and Wang [51] investigated the intermedia agenda-setting effects during China's Changsheng vaccine crisis, uncovering a two-way influence between traditional media and WeChat's we-media, where each influenced the other under specific frames. Thus, our study aims to address this gap by exploring the dynamics between Twitter and news media agendas concerning COVID-19 vaccines.

While natural language processing techniques have been extensively utilized to uncover latent topics and sentiments from vast amounts of data, a majority of studies within this domain have primarily focused on analyzing data from a single media outlet [13,15,17,52]. For instance, Lyu et al. [17] analyzed 1499,421 unique tweets originating from 583, 499 distinct users between March 11, 2020, and January 31, 2021. Their study identified five overarching themes: opinions and emotions regarding vaccines and vaccination, knowledge about vaccines and vaccination, global perspectives on vaccines, vaccine administration, and updates on vaccine development and authorization. Another study by Wang et al. [13] explored a dataset comprising 154,978 geo-tagged coronavirus tweets from March 20th to August 9th, 2020, identifying five major topics associated with COVID-19 vaccines: science, coping without vaccine, immunity boost, vaccine race, and politics around vaccine. Although the specific themes varied based on different stages of the pandemic, both studies observed that the public discourse surrounding COVID-19 vaccines on Twitter was heavily influenced by significant events related to these vaccines and aligned with active news topics in mainstream media [13,17]. However, a comprehensive statistical analysis has yet to be conducted to substantiate this influence.

In addition to examining individual media outlets, researchers have also explored multiple media outlets to offer comparative descriptive insights. For instance, de Melo and Figueiredo [12] conducted a study comparing news articles and tweets about COVID-19 in Brazil. Their findings revealed that while Twitter displayed similar topic coverage to news media, there were differences in theme distribution and entity diversity. Similarly, Hussain et al. [53] compared public attitudes towards COVID-19 vaccines on Facebook and Twitter in the United Kingdom and the United States, while Luo et al. [54] examined the semantic and sentiment aspects of COVID-19 vaccine discussions between Twitter and Weibo. However, it is worth noting that these studies primarily presented descriptive findings, lacking statistical tests to explore the dynamic agenda-setting effect, specifically how traditional media and Twitter influence each other's agenda in a public health crisis context.

Building upon the literature above, our study delved into the examination of temporal patterns in content topics and sentiment surrounding COVID-19 vaccines. Specifically, we investigated these patterns within both Twitter discussions and traditional news headlines. Additionally, we explored the intricate dynamics of agenda-setting between tweets and news headlines. To guide our research, we formulated the following research question:

Research Question 1 (RQ1): What are the differences between tweets and news headlines regarding the evolving content topics and content valence related to COVID-19 vaccines?

Research question 2 (RQ2): What are the dynamics of agenda-setting between tweets and news headlines in content topics and content valence of COVID-19 vaccines?

The results of this study will contribute to the existing body of knowledge by enhancing our comprehension of the evolutionary patterns of topics and valence concerning COVID-19 vaccines on Twitter and within two prominent newspapers in the United States. Moreover, it will extend our understanding of the dynamic agenda-setting effects within public health, specifically concerning COVID-19 vaccines.

Method

In our study, we leveraged two natural language processing techniques, namely topic modeling and sentiment analysis, to automatically identify the cognitive and affective attributes of COVID-19 vaccines, respectively. To address the cognitive attributes, we employed Latent Dirichlet Allocation (LDA) to operationalize content topics [55]. The application of LDA-based analysis has demonstrated empirical efficiency and validity in analyzing mass communication text, particularly in the context of large-scale social media data [56]. As for the affective attributes, we operationalized them as content valence, which was determined through sentiment analysis. Sentiment analysis is a dictionary-based text classification approach commonly employed to ascertain valence and identify affective emotions within a message by comparing it to a pre-calibrated sentiment lexicon [16]. This method has been utilized in recent studies to analyze the sentiments expressed in tweets related to COVID-19 vaccines (e.g., [17,53,57]).

We tested the dynamics of agenda-setting effects using Granger causality tests [19,20] on the temporal patterns of content topics and content valence. This methodology, commonly employed by econometricians, enables us to infer causality and has been widely utilized in recent agenda-setting research [39,58–60]. Examining the temporal trends in content topics and valence is crucial for comprehending the dynamics of agenda-setting concerning rapidly evolving issues such as COVID-19 vaccines. The causality inference derived from the parallel time series can yield valuable insights into how information regarding COVID-19 vaccines is reported and disseminated. Specifically, it helps identify the issues where the editorial attention of elite news media Granger-causes the level of social media attention and the issues where the influence is reversed.

Data source

We collected tweets and news headlines about COVID-19 vaccines from January 1, 2020, to April 30, 2021. The start date was chosen to reflect the beginning of the development of COVID-19 vaccines [61].

To access the public agenda as reflected on Twitter discourses, we collected publicly available original tweets about COVID-19 vaccines using snscrape [62], which were further filtered on user profile data to include only tweets in English and from US-based users. This resulted in a total of 501,531 tweets from 143,172 unique users. While acknowledging that online tweets do not equal "public opinion" in general as Twitter users are not demographically representative [63], other methods, such as public opinion polls with self-reported phone or online survey questions, could also be biased and distorted. Thus, we consider text mining from the big data perspective to represent an increasingly important "instantiation of public opinion just as the outcomes derived from those who choose to answer surveys and those who choose to vote" [60].

To access the news agenda from the elite news media, we collected 7284 news headlines about COVID-19 vaccines from The New York Times and The Washington Post, which have been conventionally considered to set the agenda of other media outlets from numerous agenda-setting studies (e.g., [18,64]). The New York Times and The Washington Post are exemplars for their extensive reach and authoritative reporting, distinguished not just by their historical significance but by their consistent role in shaping both public opinion and media landscapes. They serve as benchmark publications for news outlets

nationwide, often setting the agenda for what becomes national news; other media frequently cite them, and their stories consistently lead to policy changes, public awareness, and national conversation [9,63]. In addition, the two newspapers are chosen to cover both angles from domestic and international editorial attention on COVID-19 vaccines, given that The Washington Post is considered one of the best newspapers in domestic coverage, and The New York Times is a significant source of international information for US readers [9,64]. This approach ensures that our study encompasses a broad spectrum of narratives, thus accurately reflecting the complex nature of U.S. media discourse. Thus, the selection of the two news newspapers as representative elite news media in the U.S. not only considers their historical depth and journalistic rigor but also their diverse readership, editorial stance, and reporting styles. The decision to analyze only news headlines is grounded in the comparability of their brevity to the concise nature of tweets. During the study period, Twitter had a limitation of 280 characters, longer than its original 140 characters limitation but still "short and brief" [65].

The online news was collected using Google BigQuery on GDELT's GEO 2.0 API [66] in its gdelt-bq.covid19.onlinenewsgeo databases. GDELT is an open database that provides various content, including worldwide news. Scholars in multiple academic disciplines have used data from the GDELT database to collect and analyze news content (e.g., [39,67]). Its gdelt-bq.covid19.onlinenewsgeo database contains online news articles in the English language that have at least one of the following keywords: "social distan*," "quarantin*," "lockdown*," "stay at home," "shelter in place," "self isolat*," "*virus*," "Covid-19," and "Sars-Cov-19," where "*" indicates a character string of any length, to capture all variations of the same word.

Drawing on prior social media studies on vaccines [68,69], we developed search keywords to filter Twitter and GDELT news data by balancing general COVID-19 vaccine information, brand-specific information, and technology-specific information. For brand-specific information, we included four brands: Pfizer-BioNTech, Moderna, Johnson & Johnson/Janssen, and AstraZeneca. Unlike the first three vaccines, which were authorized for emergency use as of the end of our study period on April 30, 2021 [70], we included AstraZeneca because it too, received considerable media coverage and public attention in the US, for its conditional marketing authorizations in the European Union countries [71]. Pfizer-BioNTech and Moderna used mRNA technology, and Johnson & Johnson and AstraZeneca-Oxford used the viral vector technology.

As a result, a tweet or a news headline from the two newspapers was included if it contained the keyword "vaccine," together with one of the following keywords: "COVID," "COVID19," "COVID-19," "Pfizer," "Pfizer-BioNTech," "Moderna," "Johnson & Johnson," "Janssen," "AstraZeneca," and "Oxford-AstraZeneca"; or it contained the keyword "vaccine," together with one of the following combinations: "mRNA" & "COVID," "viral vector" & "COVID," and "adenovirus" & "COVID." Moreover, we checked government Twitter accounts such as the US CDC and FDA to explore hashtags for tweets and added tweets that contained either "#covid19vaccine" or "#covidvaccine" in our dataset.

Data processing

We combined the Twitter and news headline datasets into one final dataset. Regardless of the source, each tweet or news headline is considered a record in all later steps.

To perform topic model training, the final dataset was pre-processed through the following four steps: tokenization, stop word removal, bigram grouping, and lemmatization. More specifically, each record is first converted into a list of tokens, the smallest units for all latter steps, such as a single word or an inseparable sequence of characters like a phone number or a URL. Second, stop words, i.e., tokens which are grammatically necessary but don't add meanings, were removed, such as articles or conjunctions. We used the standard stop word library from nltk [72] and extended it to include tokens such as https. Third, some adjacent token pairs, such as New York, are inseparable to keep the original meaning. We grouped such pairs of tokens as single units, i.e., the bigrams. We trained a bigram model on our entire dataset using Gensim [73] and labeled the bigrams in the list of tokens. Finally, to reduce the complexity of language modeling, all tokens were mapped to their lemma's, the root form from which all the variations are derived, such as verb tenses. This pre-processed dataset is used as the corpus for topic modeling.

For each record, we used a pre-trained model from TextBlob [74] to calculate its valence score. The score is a floating measure from -1 and 1, which indicates the most negative and positive emotional responses, respectively, normalized from the model's training text.

Measures

Content topics

The tweets and news headlines in our final dataset were subjected to topic modeling using the LDA algorithm [55]. Such topic models can assign a probability score to each token belonging to a latent topic defined by the model, and hence map out a distribution of topics for any sequence of text, such as a tweet or a news headline. The latent topics are represented by a Dirichlet distribution of tokens [75], with a token-topic matrix of conditional probabilities for a word to appear in the topic. In our study, an LDA topic model was trained over the corpus of combined tweets and news headlines. We interpreted the latent topics by examining the token-topic matrix and reviewing the original texts from tweets and news headlines, for which a particular topic is the most predominant.

Topic salience

Topic salience was operationalized by topic size. The size of each topic on a single record can be represented as the weighted sum of the loading score from LDA over its popularity score. More specifically, the popularity score of a record k is calculated as $P_k = 1 + n_k^{like} + n_k^{share}$, where P_k is the popularity score of a tweet k, n_k^{like} is the number of likes, and n_k^{share} is the number of shares. For a tweet, the number of shares is the retweet count. For news headlines, we set both n_k^{like} and n_k^{share} to 0, where we conceptualize the attention of each news article as its editorial decisions only.

A record's topic-specific loading score from LDA, which is represented by the percentage of tokens assigned to the topic by the LDA model, is calculated as $v_{j,k} = n_{j,k} / \sum_i n_{i,k}$, summed over all tokens from the record, where $n_{j,k}$ denotes the number of tokens assigned to the topic j in the record k. It indicates the conditional probability that a topic is likely to occur in a record. The topic sizes of all loading topics in a single record add to 1. That is, $\sum_j v_{j,k} = 1$. For a sample of records, the size of each topic $V_{j,k}$ is represented by the weighted sum of individual topic sizes of all records in the sample, over its popularity P_k. More specifically, we calculate the sample topic size for tweets and news headlines as $V_j^{tweets} = \sum_k^{N_{tweets}} v_{j,k} P_k$, and $V_j^{news} = \sum_k^{N_{news}} v_{j,k} P_k$, respectively.

A normalized sample topic size is the sample topic size percentage of the sum of popularities throughout the corpus. That is,

$$\overline{V}_j^{tweets} = V_j^{tweets} \Big/ \sum_k^{N_{tweets}} P_k \text{ and } \overline{V}_j^{news} = V_j^{news} \Big/ \sum_k^{N_{news}} P_k. \quad (1)$$

Content valence

For each tweet or news headline k, we denote the valence score using TextBlob [74] by s_k. For any given sample of size N_{tweets} or N_{news}, we denote the sample valence of topic j as S_j^{tweets} and S_j^{news}, following

$$S_j^{tweets} = \sum_k^{N_{tweets}} v_{j,k} P_k s_k \Big/ V_j^{news} \text{ and } S_j^{news} = \sum_k^{N_{news}} v_{j,k} P_k s_k \Big/ V_j^{news}, \quad (2)$$

which are weighted averages of a record valence over the topic sizes and popularities, normalized to the sample topic sizes.

Data analysis

To study the changes in topic size and valence over the study period, we denote a sequence of timestamps by t_i, separated by 1-day intervals, with t_0 and t_{end} representing January 1, 2020, and April 30, 2021, respectively. Two-time series can then be constructed for each topic j on the normalized topic size $\{\overline{V}_j^{tweets}(t_i),\ t_0 \le t_i \le t_{end}\}$ and $\{\overline{V}_j^{news}(t_i),\ t_0 \le t_i \le t_{end}\}$, and topic valence $\{S_j^{tweets}(t_i),\ t_0 \le t_i \le t_{end}\}$ and $\{S_j^{news}(t_i),\ t_0 \le t_i \le t_{end}\}$, where for each time interval i, the sample is collected such that the tweets and news headlines were published between t_{i-1} and t_i.

We used Granger analysis to test the agenda-setting dynamics. In this framework, x is said to "Granger-cause" y, under the condition that y can be better predicted from past values of x and y together, than from past values of y alone [76]. Thus, if the predictability of traditional news media on Twitter discourse about the same topic in preceding days is above its autoregressive function, which is its impact from the past to the days preceding, we can conclude that the attention in traditional news media Granger-caused the level of social media attention for this topic, vice versa. In our study, we studied bi-directional Granger casualties on each pair of tweet and news headline time series for a given topic, for their normalized topic sizes and sentiment valances, respectively.

In a typical setting of the Granger causality test, two-time series are required to be stationary, or if not stationary, transformed into stationary series by taking the first (or higher) difference. Therefore, in our study, we examined the stationarity of each time series by the Augmented Dickey-Fuller test, the null hypothesis of which states a unit root exists, and thus the time series is non-stationary [77]. For time series failing to pass the Augmented Dickey-Fuller test, we recursively take the first difference in the data until the null hypothesis can be rejected, resulting in a stationary series.

To study the Granger causality between two time series, an F test is typically performed on the regression between one time series and the other with a specific time lag. The test statistically implies whether the past values of one variable offer statistically significant information to, or in other words, Granger-caused the other variable. In a typical setting, multiple values on the time lag are considered. In our study, we report the most significant statistic among time lags between 1 and 7 days, following insights from previous studies, which suggest that agenda-setting often occurs in a week or less [78,79]. Roberts et al. [78] found that the lag between traditional news and online discussion varied from 1 to 7 days, with day 7 producing the most effects.

Results

Content topics

During the model training process, we focused on optimizing our final model based on its coherent score, C_v, which evaluates the likelihood of token co-occurrence within the same topics [80]. To achieve this, we conducted a grid search on the Dirichlet concentration parameters, exploring a range of topic sizes from 3 to 21. The resulting model yielded a coherent score of C_v=0.43. The findings and interpretations of the 13 identified topics were thoroughly discussed among the research team members. To ensure accuracy, the team extensively reviewed high-frequency tokens and examined sample tweets and news headlines with high topic-specific loadings. A summary of the 13 topics is presented in Table 1.

Fig. 1 presents a comparative analysis of topic sizes between tweets and news headlines regarding COVID-19 vaccines. Within tweets, the most prominent topic, indicated by the largest topic size, revolved around people's views (Topic 1), closely followed by efficacy and rollout (Topic 2), and vaccine access (Topic 3). In contrast, news headlines predominantly covered the topic of efficacy and rollout (Topic 2), followed by research and risk (Topic 4), and people's views (Topic 1). Notably, the distribution of topic sizes among the 13 identified topics was more evenly spread among news headlines compared to tweets. Twitter discourses exhibited a stronger focus on the top five ranked topics.

Table 1
Summary of Topics.

Topic No.	Topic label	Top 15 words by frequency and relevance (λ = 0.6)
1	People's views	go, take, want, know, come, vaccine, let, would, say, think, trump, tell, right, people, thing
2	Vaccine efficacy and rollout	shoot, case, virus, death, still, variant, vaccine, spread, get, number, new, infection, protect, rate, effective
3	Vaccine access	community, vaccine, help, share, access, join, pandemic, global, effort, member, distribution, ensure, rollout, support, discuss
4	Research and risk	use, vaccine, study, safety, data, show, develop, rare, research, result, blood_clot, create, response, read, drug
5	Vaccine eligibility	dose, first, week, shot, second, received, day, administer, vaccine, trial, next, begin, morning, last, least
6	Vaccination status	get, safe, thank, do, today, back, part, great, work, way, hope, opinion, good, look, family
7	Vaccine appointment	appointment, update, site, available, vaccine, schedule, visit, information, open, sign, check, call, find, link, vaccination
8	Reopen economy	state, retweet, require, plan, school, business, local, run, travel, passport, urge, employee, return, center, pharmacy
9	Age and issues	vaccinate, get, fully, age, old, risk, people, adult, eligible, vaccinated, group, vaccine, line, worker, love
10	Resident care	health, question, resident, offer, care, tomorrow, answer, concern, medical, public, supply, official, staff, provider, area
11	Feeling and side-effect	live, feel, mask, arm, watch, side_effect, wear, please, hour, black, video, yesterday, push, nee, experience
12	children and concern	pause, increase, fact, reach, child, change, hit, due, wait, surge, list, clot, warn, ever, effect
13	Student and campaign	clinic, student, woman, Biden, bring, encourage, together, card, trust, demand, campaign, nation, meet, deliver, college

Fig. 2a and b depicted the normalized topic size over time for tweets and news headlines, respectively. Each data point represents a sample topic size aggregated over a 7-day period. The analysis reveals notable similarities and differences in the trends observed.

Both tweets and news headlines exhibited spikes in topic sizes for most topics during significant events related to COVID-19 vaccines. Notably, a spike in topic sizes occurred when Pfizer published its vaccine results on November 9, 2020 [81]. This event generated considerable public interest and discussion, likely leading to a surge in topic sizes on both platforms. Furthermore, another spike in topic sizes was observed when the FDA advisory panel endorsed the Pfizer-BioNTech COVID-19 vaccine in December 2020 [82]. These spikes in topic sizes indicate heightened engagement and public attention surrounding vaccine-related news and developments during those periods. After experiencing a slight drop in popularity following the initial spikes, the discussions regarding COVID-19 vaccines remained consistently high in 2021.

However, while news headlines relatively consistently covered all topics throughout the study period, discussions on Twitter displayed larger fluctuations. This discrepancy suggests that Twitter discussions

Source	Topic no.	Valence	Topic lable
Tweets	1	0.1538	People's view
	2	0.155441	Efficacy and rollout
	3	0.158235	Vaccine access
	4	0.142016	Research and risk
	5	0.148453	Vaccine eligibility
	6	0.143153	Vaccination status
	7	0.149479	Vaccine appointment
	12	0.12276	Children and concern
	8	0.11697	Reopen economy
	9	0.101968	Age and issues
	10	0.105989	Resident care
	11	0.095881	Feeling and side-effect
	13	0.151179	Student and campaign
News	2	0.046837	Efficacy and rollout
	4	0.035303	Research and risk
	1	0.073044	People's views
	3	0.06963	Vaccine access
	9	0.052238	Age and issues
	11	0.050746	Feeling and side-effect
	7	0.049508	Vaccine appointment
	12	0.064193	Children and concern
	5	0.06685	Vaccine eligibility
	6	0.045403	Vaccination status
	8	0.049374	Reopen economy
	10	0.04566	Resident care
	13	0.051331	Student and campaign

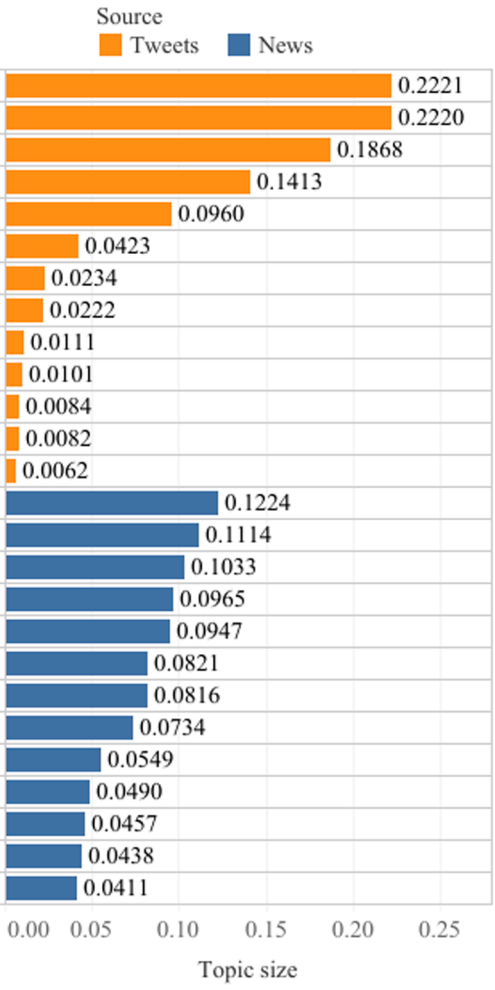


Fig. 1. Comparisons of topic sizes and average valence scores between tweets and news headlines.

are more susceptible to variations in popularity and engagement compared to news headlines. Interestingly, the discussion about vaccines appeared much later on Twitter compared to news headlines. On Twitter, vaccine-related discussions gained traction after the WHO declared COVID-19 a Pandemic on March 11, 2020 [82]. In contrast, news headlines had already started covering topics regarding the COVID-19 vaccine even before February 3, 2020, when the US declared a public health emergency [82]. This discrepancy highlights a temporal disparity between the two platforms, indicating that news headlines were more proactive in reporting vaccine-related information prior to its widespread discussion on Twitter.

Topic valence

Overall, tweets about the COVID-19 vaccine showed a more positive average valence score (0.15) than news headlines (0.05). As shown in Fig. 1, For tweets, the most positive topic was *vaccine access* (Topic 3), followed by *efficacy and rollout* (Topic 2) and *people's views* (Topic 1). For news headlines, the most positive topic was *people's views* (Topic 1), followed by *vaccine access* (Topic 3) and *vaccine eligibility* (Topic 5). While none of the topics in news headlines has an average valence score above 0.1, on a scale from -1 to 1, only one topic in Twitter discussions shows an average valence score below 0.1.

Fig. 3 shows the comparison in the change of valence for each topic over time for tweets and news headlines. Both tweets and news headlines exhibited fluctuating average valence scores for different topics, particularly during the initial period of the study, which corresponds to the first quarter of 2020. At this stage, the average valence scores significantly differed from one another for both tweets and news headlines. However, as the study period progressed, the average sentiment scores for different topics gradually converged, eventually stabilizing around two distinct levels for both tweets and news headlines.

Despite the overall similarities in the trend, the valence trends for tweets and news headlines also demonstrated noticeable differences. At the beginning of the pandemic, there was a significant drop and fluctuations in valence scores on Twitter; however, all average valence scores remained positive. In contrast, news headlines showed negative valences at multiple instances during the same period. During the later stages of the pandemic, while tweets leveled off at a slightly positive valence level, news headlines predominantly displayed neutral valence. This indicates that Twitter discussions maintained a relatively stable and positive sentiment, while news headlines remained more neutral in their coverage.

The dynamics of agenda-setting

To measure the salience transfer of content topics between tweets and headlines, we used the Granger causality analysis to examine 13 pairs of topic size time series for headline news and tweets, for each of the 485 days of our study period. Similarly, to measure the affective salience transfer of valence, we examined 13 pairs of valence time series. The results are presented in Table 2a and Table 2b, respectively.

The results revealed that for 9 out of the 13 topics, time series in topic sizes in tweets, Granger caused topic sizes in news headlines. The most

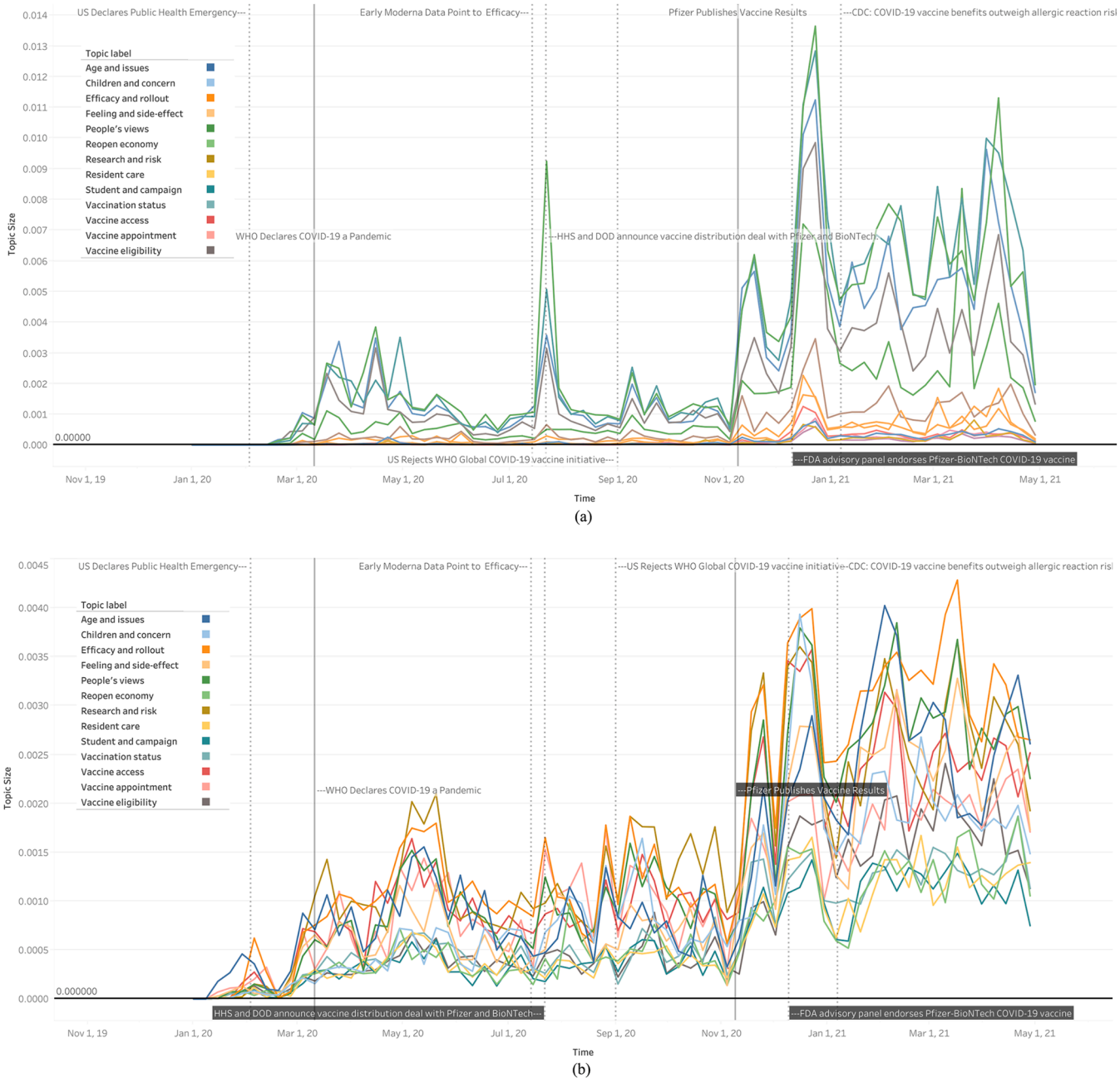


Fig. 2. (a) Normalized topic size over time of tweets for each topic. (b) Normalized topic size over time of news headlines for each topic.

significant impact of tweets on news headlines was found at different lags: 6 days on *vaccine access* (Topic 3), *research and risk* (Topic 4), *vaccine eligibility* (Topic 5), *vaccine appointment* (Topic 7) and *reopen economy* (Topic 8); 2 days on *vaccination status* (Topic 6), five days on *resident care* (Topic 10), three days on *feeling and side-effect* (Topic 11), and seven days on *children and concern* (Topic 12). On the other hand, news headlines Granger caused Twitter discourses in 9 out of the 13 topics. The difference in lags for the most significant impact was also revealed: 5 days on *people's view* (Topic 1), *efficacy and rollout* (Topic 2), *vaccine access* (Topic 3), and *vaccination status* (Topic 6); 6 days on *research and risk* (Topic 4), *age and issues* (Topic 9) and *resident care* (Topic 10); 4 days on *vaccine eligibility* (Topic 5); and seven days on *student and campaign* (Topic 13).

Tweets had a one-way influence on news headlines on four topics: *vaccine appointment* (Topic 7), *reopening the economy* (Topic 8), *feeling and side effects* (Topic 11), and *children and concern* (Topic 12). News headlines had a one-way influence on tweets on four topics: *people's views* (Topic 1), *efficacy and rollout* (Topic 2), *age and issues* (Topic 9), and *student and campaign* (Topic 13). Reciprocal causality was revealed in the remaining five topics, which were: *vaccine access* (Topic 3), *research and risk* (Topic 4), *vaccine eligibility* (Topic 5), *vaccination status* (Topic 6), and *resident care* (Topic 10).

As to salience transfer of valence between tweets and news, valence in tweets Granger caused the changes of topic valence in news headlines in 4 of the 13 tests, which were *people's views* (Topic 1), *research and risk* (Topic 4), *vaccine eligibility* (Topic 5), and *children and concern* (Topic 12). Except for Topic 1, where the lag time for most significant influence is six days, the transfer of affective tone from Twitter to news is relatively faster, with a lag time of 1 day for Topic 4 and Topic 5, and 3 days for Topic 12. On the other hand, news headlines Granger caused the changes of topic valence in tweets in 3 out of 13 topics, which were *age and issues* (Topic 9), *feeling and side-effects* (Topic 11), and *children and*

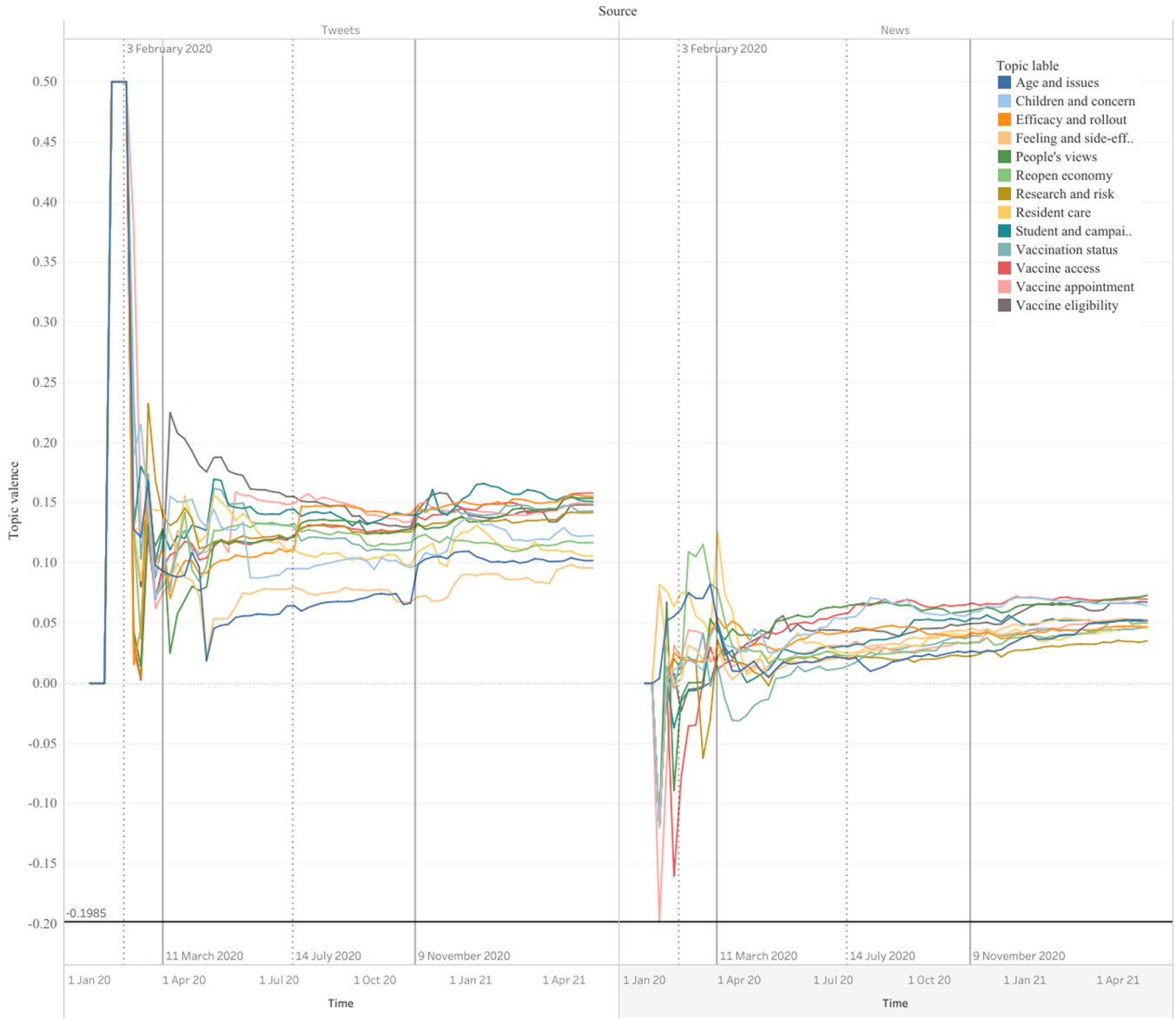


Fig. 3. Comparisons of topic valence trends between tweets and news headlines.

concern (Topic 12). The lag from news to tweets is four days for Topic 9 and Topic 11, and 2 days for Topic 12. Unlike salience transfer of content topics, mutual reciprocal causation for valence was only revealed on one topic: *children and concern* (Topic 12). It is noteworthy that the salience transfer was more evident for topic contents than for topic valence between tweets and news headlines.

Discussion

This study focused on exploring the evolving agendas of tweets and news headlines concerning the issue of COVID-19 vaccines. The analysis went beyond cognitive attributes by investigating the dynamic agenda-setting effects in terms of the transfer of both cognitive (content topics) and affective (valence) attributes. The study aimed to determine whether there was a clear dominator between elite news media and Twitter in setting the agenda for both cognitive and affective aspects of the COVID-19 vaccine issue.

The comparison of normalized topic sizes over time for tweets and news headlines reveals both similarities and differences in the observed trends. Both platforms exhibited topic size spikes during significant vaccine-related events, such as the publication of Pfizer's vaccine results and the FDA advisory panel endorsement. Additionally, sustained interest in vaccine discussions was observed in 2021. However, Twitter discussions displayed larger fluctuations in topic sizes compared to news headlines, reflecting the dynamic nature of the platform. Moreover, news headlines covered vaccine-related discussions earlier than Twitter, suggesting a temporal disparity in the reporting and emergence of vaccine-related topics between the two platforms.

The findings of this study indicate that no clear dominator emerged between elite news media and Twitter regarding agenda-setting for the cognitive and affective attributes of the COVID-19 vaccine issue. In other words, neither platform appeared to have a consistent influence in setting the agenda for the topics discussed or the emotional tone surrounding COVID-19 vaccines.

On a cognitive level, the reciprocal influence between elite news and Twitter discourse was observed for five out of the 13 topics analyzed. Elite news media was found to have an exclusive impact on Twitter discourses regarding people's views (Topic 1), efficacy and rollout (Topic 2), age and related issues (Topic 9), and student and campaign (Topic 13). Conversely, Twitter discourse exhibited greater influence on topics such as vaccine appointment (Topic 7), reopening the economy (Topic 8), feelings and side-effects (Topic 11), and children and concerns

Table 2a
Granger analysis of topic size trends in the tweets and news headlines from January 1, 2020, to April 30, 2021.

	Tweets to news headlines			News headlines to Tweets		
Topic Vectors	Chi-Square	Lag time	*P*-value	Chi-Square	Lag time	*P*-value
1 People's views	5.75	2 days	.056	24.53**	5 days	<0.001
2 Efficacy and rollout	9.13	7 days	.244	22.40**	5 days	<0.001
3 Vaccine access	26.43**	6 days	<0.001	29.41**	5 days	<0.001
4 Research and risk	19.02**	6 days	.004	15.05*	6 days	.02
5 Vaccine eligibility	16.64*	6 days	.011	15.92**	4 days	.003
6 Vaccination status	20.93**	2 days	<0.001	15.72**	5 days	.008
7 Vaccine appointment	25.52**	6 days	<0.001	12.31	6 days	.055
8 Reopen economy	18.18**	6 days	.006	11.32	7 days	.125
9 Age and issues	11.72	7 days	.110	16.65*	6 days	.011
10 Resident care	21.51**	5 days	.001	15.11*	6 days	.019
11 Feeling and side-effect	9.23*	3 days	.026	1.72	1 days	.189
12 Children and concern	17.83*	7 days	.013	7.59	3 days	.055
13 Student and campaign	4.96	2 days	.084	15.02*	7 days	.036

*P <0.05.
** P <0.01.

Table 2b
Granger analysis of topic valence trends in the tweets and news headlines from January 1, 2020, to April 30, 2021.

	Tweets to news headlines			News headlines to tweets		
Topic Vectors	Chi-Square	Lag time	*P*-value	Chi-Square	Lag time	*P*-value
1 People's view	20.87**	6 days	.002	1.38	1 days	.240
2 Efficacy and rollout	1.95	1 days	.162	2.29	2 days	.318
3 Vaccine access	1.49	1 days	.222	5.88	6 days	.436
4 Research and risk	9.67**	1 day	.002	9.98	6 days	.125
5 Vaccine eligibility	10.35**	1 day	.001	0.61	1 days	.433
6 Vaccination status	8.10	5 days	.151	1.68	1 days	.195
7 Vaccine appointment	3.19	4 days	.526	3.80	2 days	.150
8 Reopen economy	2.83	2 days	.242	4.98	2 days	.083
9 Age and issues	3.30	4 days	.509	13.91**	4 days	.008
10 Resident care	7.08	5 days	.215	9.17	7 days	.241
11 Feeling and side-effect	4.56	3 days	.207	9.55*	4 days	.049
12 Children and concern	7.94*	3 days	.047	7.77*	2 days	.021
13 Student and campaign	7.96	7 days	.336	6.21	7 days	.515

*P <0.05. **P <0.01.

(Topic 12). These findings align with previous research on the dynamic relationship between traditional media and social media, which lacks a clear dominator [60]. Traditional news and social media possess their unique resonances and interactions, exchanging insights on public attention [5,25,27,83], with their responses intertwined based on specific issues and their attributes.

In our analysis, elite news demonstrated a greater influence on topics that hold societal significance within the context of COVID-19 vaccination, such as efficacy and rollout, age-related issues, and people's views. This finding aligns with previous research that emphasizes the audience's need for orientation, enhancing agenda-setting effects [40]. Particularly when individuals lack direct and personal experiences, they heavily rely on media coverage to comprehend unobtrusive issues, which is particularly relevant during a public health crisis like the COVID-19 pandemic. Therefore, echoing previous studies, our findings highlight the continued significance of elite media in setting the agenda for public discourse on COVID-19 vaccines.

Conversely, our study also suggests that individuals on Twitter, play a significant role in shaping the COVID-19 agenda, sometimes independently of traditional media's agenda-setting. As observed in our results, news media exhibited insignificant unidirectional Granger causality on nine topics discussed on Twitter. Contrarywise, four out of these nine topics demonstrated influential agenda-setting effects from Twitter to traditional news, including vaccine appointment (Topic 7), feelings and side-effects (Topic 11), children and concerns (Topic 12), and reopening the economy (Topic 8). This pattern highlights that conversations on Twitter about personal impact and emotions can set the agenda for traditional media coverage. The topic of reopening the economy, in particular, showed a unique trend where Twitter users were engaging with economic and governmental issues more than expected [60], likely due to the direct financial impact of the pandemic on individuals. The study underscores the dynamic and context-dependent relationship between social media discourse and traditional news coverage. It also emphasizes the potential for public health officials to use Twitter as a real-time monitor for public sentiment and concerns, which can inform public health strategies and communication. Twitter's responsiveness to issues of individual concern suggests it could be a valuable channel for disseminating health messages that resonate on a personal level, such as information about vaccine appointments or responses to parental concerns regarding vaccinations for children.

Furthermore, our findings shed light on the effective lag time for the transfer of topics between tweets and news headlines. The analysis revealed that the most common effective lag time for the transfer of topics from tweets to news headlines was six days, encompassing a total of 5 topics. The longest effective lag time observed was seven days, as evident in Topic 12, while the shortest lag time was two days, exemplified by Topic 6. Conversely, the most common effective lag time for the transfer of topics from news headlines to tweets was five days, encompassing four topics, followed closely by six days, observed for three topics. The longest effective lag time between news headlines and tweets was seven days, as observed in Topic 13, whereas the shortest lag time was four days, exemplified by Topic 5.

These findings support previous observations suggesting that agenda-setting often occurs within a week or less [78,79]. Moreover, the lag between traditional news and online discussion has been reported to vary from 1 to 7 days [78]. The consistent patterns identified in the effective lag times between tweets and news headlines further underscore the temporal dynamics involved in the agenda-setting process. The transfer of topics from one platform to the other is not instantaneous but takes time for the influence to manifest.

Moving to the affective level, our analysis revealed fewer instances of Granger causations in valence trends compared to cognitive attributes. Furthermore, when the transfer of valence occurred from tweets to news headlines, the affective lag time was relatively faster, with examples such as one day observed for Topic 4 and Topic 5. In contrast to the transfer of cognitive attributes indexed by topic, the transfer of affective attributes indexed by valence demonstrated weaker patterns. This discrepancy can be attributed, in large part, to the neutral valence typically associated with news headlines, as the pursuit of objectivity has long been a fundamental principle in news coverage.

In summary, our study provides an in-depth examination of how elite news media and social media platforms, especially Twitter, reciprocally influence the public conversation surrounding COVID-19 vaccines. It enriches agenda-setting theory by uncovering the complex, multifaceted interactions between these media in shaping both the cognitive and emotional dimensions of the vaccine discourse. Consistent with prior research, we find that elite media outlets like The New York Times and The Washington Post significantly influence healthcare narratives. However, our study goes further, examining the second level of agenda-setting, which allows us to explore not only the topics themselves but also the public's perception and emotional engagement with these topics. Our analysis shows that COVID-19 vaccination is a complex and multifaceted issue. Elite media tends to have a greater impact on topics of societal importance related to COVID-19 vaccination, whereas Twitter discussions, particularly those focusing on personal impact and emotions, often influence the coverage of these topics in traditional media. By applying the agenda-setting theory within the public health context, we highlight the critical role of both cognitive and emotional elements in shaping public perception and response. These findings offer valuable guidance for public health officials and communication strategists in crafting informative and emotionally resonant messages and utilizing communication channels effectively. Understanding these agenda-setting dynamics is increasingly vital in the evolving media landscape for effective public health communication, particularly in times of crisis.

Limitations

This study employed computational methods to identify cognitive and affective attributes related to the COVID-19 vaccine issue across Twitter discourse and two elite newspapers. Granger causality tests were subsequently employed to quantify the dynamic agenda-setting effects by examining the time series pairs of topic sizes and their corresponding valences between the two platforms. However, it is important to acknowledge certain limitations in interpreting the Granger causality tests.

While Granger causality tests suggest a causal relationship, it is crucial to note that the observed findings should not be construed as indicating that Twitter discourse "causes" traditional news topic attention, or vice versa. It is possible that traditional news and Twitter operate based on their unique characteristics and responsiveness to global events, rather than one directly influencing the other in a causal manner [60].

Furthermore, the generalizability of the findings is limited by the focus on the English language and the inclusion of US-based newspapers and tweets. The dynamics of agenda-setting and the relationship between platforms and news sources may differ in other languages and cultural contexts. Including data from a broader range of countries and languages would enhance the external validity of the findings.

Additionally, the computational methods used to analyze cognitive and affective attributes rely on algorithms and models, which may have inherent limitations and biases. Likewise, the operationalization of content valence as a floating score ranging from -1 to 1 has limitations when studying the impact of specific emotions on dynamic agenda-setting effects. It does not allow for a nuanced examination of how particular emotions, such as fear expressed on Twitter, influence subsequent news reporting or vice versa. Future studies could explore the influence of specific emotions on agenda-setting processes in the context of public health.

It is also important to acknowledge that the COVID-19 pandemic and related discussions are constantly evolving. The study's findings are based on a specific time period and may not capture the entirety of the discourse or reflect subsequent developments or shifts in public opinion. Similarly, the sampling rate of every 24 hours in the time series may have limited the ability to capture more granular phenomena occurring within a single day. More frequent sampling intervals could provide a more detailed understanding of agenda-setting dynamics.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used ChatGPT in order to improve readability. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

[1] P. Ball, The lightning-fast quest for COVID vaccines - and what it means for other diseases, Nat. News (2020). https://www.nature.com/articles/d41586-020-03626-1.
[2] C.L. Coleman, The influence of mass media and interpersonal communication on societal and personal risk judgments, Commun. Res. 20 (4) (1993) 611–628, https://doi.org/10.1177/009365093020004006.
[3] C.A. Lin, C. Lagoe, Effects of news media and interpersonal interactions on H1N1 risk perception and vaccination intent, Commun. Res. Rep. 30 (2) (2013) 127–136, https://doi.org/10.1080/08824096.2012.762907.
[4] Y. Zhu, R. Wei, V. Lo, M. Zhang, Z. Li, Collectivism and altruistic behavior: a third-person effect study of COVID-19 news among wuhan residents, Glob. Media China 6 (4) (2021) 1–16, https://doi.org/10.1177/2059436421104556.
[5] M.E. McCombs, D.L. Shaw, The agenda-setting function of mass media, Public Opin. Q. 36 (2) (1972) 176, https://doi.org/10.1086/267990.
[6] M.E. McCombs, S. Valenzuela, Setting the Agenda: The Mass Media and Public Opinion, Polity, 2021.
[7] M. Broersma, T. Graham, Twitter as a news source, Journal. Pract. 7 (4) (2013) 446–464, https://doi.org/10.1080/17512786.2013.802481.
[8] S.J. Moon, P. Hadley, Routinizing a new technology in the newsroom: Twitter as a news source in mainstream media, J. Broadcast. Electron. Media 58 (2) (2014) 289–305, https://doi.org/10.1080/08838151.2014.906435.
[9] Y. Su, P. Borah, Who is the agenda setter? examining the intermedia agenda-setting effect between Twitter and newspapers, J. Inf. Technol. Politics 16 (3) (2019) 236–249, https://doi.org/10.1080/19331681.2019.1641451.
[10] S. Valenzuela, S. Puente, P.M. Flores, Comparing disaster news on Twitter and television: an intermedia agenda-setting perspective, J. Broadcast. Electron. Media 61 (4) (2017) 615–637, https://doi.org/10.1080/08838151.2017.1344673.
[11] M. McCombs, J.P. Llamas, E. Lopez-Escobar, F. Rey, Candidate images in Spanish elections: second-level agenda-setting effects, Journal. Mass Commun. Q. 74 (1997) 703–717, https://doi.org/10.1177/107769909707400404.
[12] T. de Melo, C.M. Figueiredo, Comparing news articles and tweets about COVID-19 in Brazil: Sentiment analysis and Topic modeling approach, JMIR Public Health Surveill. 7 (2) (2021), https://doi.org/10.2196/24585.
[13] Y. Wang, M. Shi, J. Zhang, What public health campaigns can learn from people's Twitter reactions on mask-wearing and covid-19 vaccines: A topic modeling approach, Cogent. Soc. Sci. 7 (1) (2021), https://doi.org/10.1080/23311886.2021.1959728.
[14] Z. Xu, L. Ellis, M. Laffidy, News frames and news exposure predicting flu vaccination uptake: Evidence from U.S. newspapers, 2011–2018 using computational methods, Health Commun. 37 (1) (2020) 74–82, https://doi.org/10.1080/10410236.2020.1818958.
[15] J. Zhang, Y. Wang, M. Shi, X. Wang, Factors driving the popularity and virality of COVID-19 vaccine discourse on Twitter: Text mining and data visualization study, JMIR Public Health Surveill. 7 (12) (2021), https://doi.org/10.2196/32814.
[16] B. Liu, Sentiment analysis and opinion mining, Synth. Lect. Hum. Lang. Technol. 5 (1) (2012) 1–167, https://doi.org/10.2200/s00416ed1v01y201204hlt016.
[17] J.C. Lyu, E.L. Han, G.K. Luli, COVID-19 vaccine-related discussion on Twitter: topic modeling and sentiment analysis, J. Med. Internet Res. 23 (6) (2021) e24435, https://doi.org/10.2196/24435.
[18] P.J. Shoemaker, S. Reese, L. Danielian, Intermedia influence and the drug issue: converging on cocaine. Communication Campaigns About Drugs: Government, Media, and the Public, Routledge, 2011, pp. 29–46.
[19] C.W. Granger, Investigating causal relations by econometric models and cross-spectral methods, Econometrica 37 (3) (1969) 424, https://doi.org/10.2307/1912791.

[20] N. Wiener, The theory of prediction, F. E. Beckenbach & R. Weller. Modern Mathematics For Engineers, McGraw Hill, 1956.
[21] P.J. Shoemaker, Mediating the Message in the 21st Century: A Media Sociology Perspective, Routledge, 2014.
[22] W. Wanta, J. Foote, The president-news media relationship: a time series analysis of agenda-setting, J. Broadcast. Electron. Media 38 (4) (1994) 437–448, https://doi.org/10.1080/08838159409364277.
[23] S.D. Reese, L.H. Danielian, Intermedia influence and the drug issue: converging on cocaine, Shoemaker P. J.. Communication Campaigns About drugs: Government, media, and the Public, Lawrence Erlbaum, New York, NY, 1989, pp. 29–46.
[24] K.D. Sweetser, G.J. Golan, W. Wanta, Intermedia agenda-setting in television, advertising, and blogs during the 2004 election, Mass Commun. Soc. 11 (2) (2008) 197–216, https://doi.org/10.1080/15205430701590267.
[25] P.J. Boczkowski, News At Work: Imitation in an Age of Information Abundance, The University of Chicago Press, 2011.
[26] F. Hanusch, D. Nölleke, Journalistic homophily on social media, Digit. Journal. 7 (1) (2019) 22–44, https://doi.org/10.1080/21670811.2018.1436977.
[27] E. Lopez-Escobar, J.P. Llamas, M. McCombs, F.R. Lennon, Two levels of agenda-setting among advertising and news in the 1995 Spanish elections, Polit. Commun. 15 (2) (1998) 225–238, https://doi.org/10.1080/10584609809342367.
[28] A. Delwiche, Agenda–setting, opinion leadership, and the world of weblogs, First Monday (2005), https://doi.org/10.5210/fm.v10i12.1300.
[29] J. Groshek, M. Groshek, Agenda trending: reciprocity and the predictive capacity of social networking sites in intermedia agenda setting across topics over time, Media Commun. 1 (1) (2013) 15–27, https://doi.org/10.17645/mac.v1i1.71.
[30] B. Landis, W. Allen, Cascading activation revisited: how audiences contribute to news agendas using social media, Digit. Journal. 10 (4) (2022) 537–555, https://doi.org/10.1080/21670811.2021.1962728.
[31] J.K. Lee, The effect of the internet on the homogeneity of the media agenda: a test of the fragmentation thesis, Journal. Mass Commun. Q. 84 (4) (2007) 745–760, https://doi.org/10.1177/107769900708400406.
[32] S. Meraz, Using time-series analysis to measure intermedia agenda-setting influence in traditional media and political blog networks, Journal. Mass Commun. Q. 88 (1) (2011) 176–194, https://doi.org/10.1177/107769901108800110.
[33] B. Sayre, L. Bode, D. Shah, D. Wilcox, C. Shah, Agenda setting in a digital age: tracking attention to California proposition 8 in social media, online news, and conventional news, Policy Internet 2 (2) (2010) 7–32, https://doi.org/10.2202/1944-2866.1040.
[34] C.J. Vargo, L. Guo, Networks, big data, and intermedia agenda setting: an analysis of traditional, partisan, and emerging online U.S. news, Journal. Mass Commun. Q. 94 (4) (2017) 1031–1055, https://doi.org/10.1177/1077699016679976.
[35] C.J. Vargo, E. Basilaia, D.L. Shaw, Event versus issue: Twitter reflections of major news: a case study, in: Communication and Information Technologies Annual, 9, Emerald Group, Bingley, England, 2015, pp. 215–239.
[36] M.E. McCombs, Setting the Agenda: The Mass Media and Public Opinion, John Wiley, Cambridge, UK, 2013.
[37] M. Maurer, T. Holbach, Taking online search queries as an indicator of the public agenda: the role of public uncertainty, Journal. Mass Commun. Q. 93 (2016) 572–586, https://doi.org/10.1177/1077699015610072.
[38] G. Golan, W. Wanta, Second-level agenda-setting in the New Hampshire primary: A comparison of coverage in three newspapers and public perceptions of candidates, Journal. Mass Commun. Q. 78 (2) (2001) 247–259, https://doi.org/10.1177/107769900107800.
[39] L. Guo, C. Vargo, Fake news" and emerging online media ecosystem: an integrated intermedia agenda-setting analysis of the 2016U.S. presidential election, Commun. Res. 47 (2) (2018) 178–200, https://doi.org/10.1177/0093650218777177.
[40] K. Wahl-Jorgensen, T. Hanitzsch, R. Coleman, M. McCombs, D. Shaw, D. Weaver, Agenda setting, Wahl-Jorgensen, K. & Hanitzsch, T.. The Handbook of Journalism Studies, 1st ed., Routledge, 2009.
[41] L. Guo, The application of social network analysis in agenda setting research: a methodological exploration, J. Broadcast. Electron. Media 56 (4) (2012) 616–631, https://doi.org/10.1080/08838151.2012.732148.
[42] L. Guo, M.E. McCombs, The Power of Information Networks: New directions For Agenda-Setting, Routledge, 2015.
[43] K. Welbers, Gatekeeping in the Digital Age, Vrije Universiteit, 2016.
[44] Y. Su, X. Xiao, Mapping the intermedia agenda setting (IAS) literature: current trajectories and future directions, Agenda Setting J. (2021) 1–28, https://doi.org/10.1075/ASJ.20001.SU.
[45] A. Abd-Alrazaq, D. Alhuwail, M. Househ, M. Hamdi, Z. Shah, Top concerns of tweeters during the COVID-19 pandemic: infoveillance study, J. Med. Internet Res. 22 (4) (2020) e19016, https://doi.org/10.2196/19016.
[46] C. Chew, G. Eysenbach, Pandemics in the age of Twitter: content analysis of tweets during the 2009 H1N1 outbreak, PLoS One 5 (11) (2010) e14118, https://doi.org/10.1371/journal.pone.0014118.
[47] M. Odlum, S. Yoon, What can we learn about the Ebola outbreak from tweets? Am. J. Infect. Control 43 (6) (2015) 563–571, https://doi.org/10.1016/j.ajic.2015.02.023.
[48] M.J. Paul, M. Dredze, You are what you tweet: analyzing Twitter for public health, in: Proceedings of the International AAAI Conference on Web and Social Media 5, 2011, pp. 265–272, https://doi.org/10.1609/icwsm.v5i1.14137.
[49] J. Keim-Malpass, E.M. Mitchell, E. Sun, C. Kennedy, Using twitter to understand public perceptions regarding the #HPV vaccine: opportunities for public health nurses to engage in social marketing, Public Health Nurs. 34 (4) (2017) 316–323, https://doi.org/10.1111/phn.12318 (Boston, Mass.).
[50] Y. Su, J. Hu, D.K.L. Lee, Delineating the transnational network agenda-setting model of mainstream newspapers and Twitter: a machine-learning approach, Journal. Stud. 21 (15) (2020) 2113–2134, https://doi.org/10.1080/1461670X.2020.1812421.
[51] J. Shi, H. Wang, Examining the intermedia agenda setting effects amid the Changsheng Vaccine Crisis: a computational approach, Int. J. Environ. Res. Public Health 20 (5) (2023) 4052, https://doi.org/10.3390/ijerph20054052.
[52] W. Xie, A. Damiano, C.H. Jong, Emotional appeals and social support in organizational YouTube videos during COVID-19, Telemat. Inf. Rep. 8 (2022) 100028, https://doi.org/10.1016/j.teler.2022.100028.
[53] A. Hussain, A. Tahir, Z. Hussain, Z. Sheikh, M. Gogate, K. Dashtipour, A. Ali, A. Sheikh, Artificial intelligence–enabled analysis of public attitudes on Facebook and Twitter toward COVID-19 vaccines in the United Kingdom and the United States: observational study, J. Med. Internet Res. 23 (4) (2021), https://doi.org/10.2196/26627.
[54] C. Luo, A. Chen, B. Cui, W. Liao, Exploring public perceptions of the COVID-19 vaccine online from a cultural perspective: semantic network analysis of two social media platforms in the United States and China, Telemat. Inform. 65 (2021) 101712, https://doi.org/10.1016/j.tele.2021.101712.
[55] D.M. Blei, A.Y. Ng, M.I. Jordan, Latent dirichlet allocation, J. Mach. Learn. Res. 3 (2003) 993–1022, https://doi.org/10.5555/944919.944937.
[56] L. Guo, C.J. Vargo, Z. Pan, W. Ding, P. Ishwar, Big social data analytics in journalism and mass communication, Journal. Mass Commun. Q. 93 (2) (2016) 332–359, https://doi.org/10.1177/1077699016639231.
[57] S.W. Kwok, S.K. Vadde, G. Wang, Tweet topics and sentiments relating to COVID-19 vaccination among Australian Twitter users: machine learning analysis, J. Med. Internet Res. 23 (5) (2021) e26953, https://doi.org/10.2196/26953.
[58] H.B. Brosius, H. Kepplinger, The agenda-setting function of television news, Commun. Res. 17 (2) (1990) 183–211, https://doi.org/10.1177/009365090017002003.
[59] L. Guo, Media agenda diversity and intermedia agenda setting in a controlled media environment: a computational analysis of China's online news, Journal. Stud. 20 (16) (2019) 2460–2477, https://doi.org/10.1080/1461670x.2019.1601029.
[60] W. Neuman, L. Guggenheim, S. Jang, S.Y. Bae, The dynamics of public attention: agenda-setting theory meets big data, J. Commun. 64 (2) (2014) 193–214, https://doi.org/10.1111/jcom.12088.
[61] C. Zimmer, J. Corum, S.L Wee, M. Kristoffersen, Coronavirus Vaccine Tracker, June 10, The New York Times, 2020, https://www.nytimes.com/interactive/2020/science/coronavirus-vaccine-tracker.html.
[62] Snscrape (2021, May 1). Github. Retrieved May 1, 2021, from https://github.com/JustAnotherArchivist/snscrape.
[63] A. Hughes, S. Wojcik, Ten Facts About Americans and Twitter, September 16, Pew Research Center, 2020, https://www.pewresearch.org/fact-tank/2019/08/02/10-facts-about-americans-and-twitter/.
[64] T.P. Boyle, Intermedia agenda-setting in the 1996 presidential election, J. Mass Commun. Q. 78 (1) (2001) 26–44, https://doi.org/10.1177/107769900107800103.
[65] Twitter. (2021). Counting characters. https://developer.twitter.com/en/docs/counting-characters.
[66] K.H. Leetaru, Mining libraries: lessons learned from 20 years of massive computing on the world's information, Inf. Serv. Use 35 (1-2) (2015) 31–50, https://doi.org/10.3233/ISU-150767.
[67] J. Hammond, N.B. Weidmann, Using machine-coded event data for the micro-level study of political violence, Res. Politics 1 (2) (2014), https://doi.org/10.1177/2053168014539924.
[68] P.M. Massey, M.D. Kearney, M.K. Hauer, P. Selvan, E. Koku, A.E. Leader, Dimensions of misinformation about the HPV vaccine on Instagram: content and network analysis of social media characteristics, J. Med. Internet Res. 22 (12) (2020), https://doi.org/10.2196/21451.
[69] P.M. Massey, A. Leader, E. Yom-Tov, A. Budenz, K. Fisher, A.C. Klassen, Applying multiple data collection tools to quantify human papillomavirus vaccine communication on Twitter, J. Med. Internet Res. 18 (12) (2016) e318, https://doi.org/10.2196/jmir.6670.
[70] CDC. (2023). Science brief: Covid-19 Vaccines and Vaccination. https://www.cdc.gov/coronavirus/2019-ncov/science/science-briefs/fully-vaccinated-people.html.
[71] European Commission. (2022). Safe COVID-19 Vaccines For Europeans. https://ec.europa.eu/info/live-work-travel-eu/coronavirus-response/safe-covid-19-vaccines-europeans_en.
[72] E. Loper, S. Bird, NLTK, in: Proceedings of the ACL-02 Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics, ETMTNLP, 2002, https://doi.org/10.3115/1118108.1118117.
[73] R. Řehůřek, P. Sojka, Genism–statistical semantics in Python. Poster, in: Proceedings of the 4th European meeting on Python in Science (EuroScipy, 2011), Paris, France, 2010. August 25-28, https://www.fi.muni.cz/usr/sojka/posters/rehurek-sojka-scipy2011.pdf.
[74] TextBlob. (2021). TexBlob: Simplified text processing. https://textblob.readthedocs.io/en/dev/.
[75] Allahyari, M., Pouriyeh, S., Assefi, M., Safaei, S., Trippe, E.D., Gutierrez, J.B., & Kochut, K. (2017, July 28). A brief survey of text mining: classification, clustering, and extraction techniques. arXiv.org. 10.48550/arXiv.1707.02919.
[76] J.R. Freeman, Granger causality and the times series analysis of political relationships, Am. J. Pol. Sci. 27 (2) (1983) 327–358, https://doi.org/10.2307/2111021.
[77] D.A. Dickey, W.A. Fuller, Distribution of the estimators for autoregressive time series with a unit root, J. Am. Stat. Assoc. 74 (366a) (1979) 427–431, https://doi.org/10.1080/01621459.1979.10482531.

[78] M. Roberts, W. Wanta, T.H.D. Dzwo, Agenda setting and issue salience online, Commun. Res. 29 (4) (2002) 452–465, https://doi.org/10.1177/009365020202900.
[79] W. Wanta, The Public and the National Agenda: How People Learn about Important Issues, Lawrence Erlbaum Associates, 1997.
[80] M. Röder, A. Both, A. Hinneburg, Exploring the space of topic coherence measures, in: Proceedings of the 8th ACM International Conference on Web Search and Data Mining (WSDM '15), New York, NY, USA, Association for Computing Machinery, 2015, pp. 399–408, https://doi.org/10.1145/2684822.2685324.
[81] Pfizer (2020, November 09) Pfizer and BioNTech announce vaccine candidate against COVID-19 achieved success in first interim analysis from Phase 3 Study. https://www.pfizer.com/news/press-release/press-release-detail/pfizer-and-biontech-announce-vaccine-candidate-against.
[82] CDC. (2023). CDC Museum COVID-19 Timeline. https://www.cdc.gov/museum/timeline/covid19.html.
[83] P. King, D. Shaw, D. Weaver, The press, candidate images, and voter perceptions, M. McCombs. Communication and Democracy: Exploring the Intellectual Frontiers in Agenda-Setting Theory, Lawrence Erlbaum Associates, 1997.