When agenda setting research meets big data: A literature review since 2016

Pu Yan[1], Huiyi Lyu[2], Buddy Janssen[3], Niloofar Hashemi[4] and Scott Hale[5]

Abstract

Media agendas, how they are shaped, and how they interact with public agendas are important research topics in the field of communication science. Traditionally, researchers have empirically measured media agendas using classical social science methods such as surveys or content analysis, but recent years have witnessed the rise of digital platforms becoming new arenas of public and news agendas. Meanwhile, public attention is increasingly driven not only by agendas shaped by established media agencies but also agendas emerging on social media platforms. Consequently the field of agenda setting research is changing to use large-scale, user-generated data collected from digital platforms to understand public attention and agendas as well as new methods such as Natural Language Processing (NLP) and Social Network Analysis (SNA). In this literature review, we systematically reviewed empirical research published between 2016 and 2022, aiming to explore *what are the main focuses of media agenda setting research during this period* and *how have computational methods and toolkits informed the study of media agendas at large scale.* Findings from this literature review can suggest research opportunities and challenges faced by social science researchers when adopting big data and computational methods in communication research and help outline theory development of agenda setting theory in the era of 'big data.'

Introduction

The news is increasingly being produced and read digitally, and machine learning and natural language processing algorithms are able to make every better sense of this data. It remains unclear, however, just how researchers studying agenda setting are using these new data sources and tools. We address this gap through a systematic literature review focused on the data and methods used to measure and compare media agendas.

McCombs and Shaw (1972)'s seminal study marked the start of the intellectual and research exploration of who sets the media and public agendas. They qualitatively coded news content and surveyed the local population in the lead up to the 1968 US presidential election. They found a strong, lagged correlation between issues covered on the nightly news and the importance ascribed to those issues by their survey respondents. This ability of the mass media to shape what issues are salience for the public opinion is referred to as **the first level** of agenda setting. McCombs (1977) later extended agenda setting theory by developing the concept of an attribute agenda. They theorized that the media not only shape the issues that are salient for the public, but what sub-attributes of each issue are important. This is referred to as **second-level agenda setting**, and there is active debate around how second-level agenda setting overlaps with framing and priming (Scheufele and Tewksbury 2007; Weaver 2007).

50 years after the study was first published, agenda setting theory has been widely used by social scientists in various disciplines to understand the interactions between the public and media—in particular, the transfer of salient issues between the public and media—in the marketplace of attention (McCombs and Shaw 1993). An agenda is defined as "the set of issues that are seriously considered in a policy" (Cobb et al. 1976; Kingdon and Stano 1984; Princen 2007), picking up on Cohen's idea that the media tells its

[1] Department of Information Management, Peking University
[2] School of Social Science, Tsinghua University
[3] Policy Office at European Commission
[4] School of Governance, University of Tehran
[5] Oxford Internet Institute, University of Oxford

audience "not what to think, but what to think about" (Cohen 1963).

Agenda setting theory in its traditional form suggests (i) that individuals perceive current events and issues to the largest extent through mass media, and (ii) that the media cannot report on everything equally and thus must implement selection and prioritization processes as to what is reported and the extent to which it is reported (Shoemaker and Vos 2009). This leads to the media providing their audience (iii) with an issue agenda prioritized by importance, which individuals adopt depending on their uncertainty about an issue and its personal relevance (McCombs 2005). The issues with highest priorities in the media are then most likely to be given serious consideration by the public at large.

Agenda setting research in the digital era

The proliferation of the Internet and social media further have opened up new opportunities for agenda setting by increasing the number of diverse information sources. Neuman el al. (2014) used big data to explore attention and framing in the traditional and social media for 29 political issues during 2012. They find agenda setting for these issues is not a one-way pattern from traditional media to a mass audience, but rather a complex and dynamic interaction between new and traditional media. The changing media environment has led to opportunities for theory development for agenda setting scholars. Reverse agenda setting and network agenda setting are proposed to describe and theorize media agendas in the digital era (Tran 2014).

Reverse agenda setting describes how new information and communication technologies enable direct communication between the public and media, allowing mass media to possibly be influenced by the public, especially by social media outlets (Weimann and Brosius 2017). Guo and McCombs (2011) proposed the **third-level of agenda setting** known as network agenda setting. This model assumes that media agendas follow a network-like structure instead of a linear one and the media agendas can interactively shape how the public understand current affairs.

While new theories were proposed by extending agenda setting framework to digital media research, methods advancements in social sciences, particularly new computational social science approaches (Lazer et al. 2009) have been introduced to the study of media agendas. Nonetheless, most of applications of computational approaches in social science research predate the use of any deep machine learning models or pre-trained language models. Natural Language Processing has been characterized as having three waves, with the latest wave focused on deep learning starting 2010 (Deng and Liu 2018). Nonetheless, the field of natural language processing greatly changed with the publication of BERT, a large pre-trained language model, in 2018 (Devlin et al. 2018). Methodological advancement in the field of computational social science not only provided new research opportunities for agenda setting research, but also created challenges for agenda setting scholars to adopt computational tool-kits to the theory-building of agenda setting framework.

Meanwhile, the use of big data approaches is not the ultimate solution for measuring media agendas, as McCombs et al. (2014) pointed out, media agendas on digital platforms cannot represent the whole of public opinions on social issues. This suggests that new computational approaches need to be adapted to agenda setting research not as a replacement of traditional social science methods, but rather as a complementary perspective.

Literature reviews on media agenda setting

Researchers have reviewed empirical studies on media agendas, examining the theoretical and methodological trends of media agenda research over the years. Rogers et al. (1993) reviewed the history of agenda setting research from an intellectual history perspective, by covering agenda setting publications from 1922 to 1992. The majority of agenda setting research were published after 1971, after McCombs and Shaw (1972)'s seminal study on the role of media on US election. They highlighted that the combination of the content analysis of media agendas with a survey of public opinions has been an empirical tradition since the early stage of agenda setting research.

The rise of the Internet has created new research field for media agenda scholars. Tran (2013) reviewed online agenda setting literature published over a decade (2000–2010) and found that the Internet has challenged the agenda-setting power of mainstream media while also shorten the time span of the salience of media agendas. The author also reviewed the rationalization of salience in empirical research and identified the frequency of issues or keywords to be important measurements of salience. Meanwhile, public salience could be measured via online search volume (i.e., Google Trends). The transmission of media agendas between media agencies and the public is measured mostly by correlations, but also sometimes by time series analysis. However, the lack of experiment designs might lead to difficulties in inferring any causal relationships. In the same vein, Weimann and Brosius (2017) also reviewed the shift of research directions of agenda-setting research in the online era. In addition to summarizing literature on

the inter-media agenda flows between social media and traditional media (Russell Neuman et al. 2014), the authors also noticed the complex dynamics between different types of media in setting agendas for the public. For example, some agendas went through an "agenda-cutting" process where newsworthy stories were under-reported on mainstream or social media. This suggest that the online media environment has potential to facilitate a more complicated and dynamic interactions across new media, mainstream media, and the public.

Kim et al. (2017) examined theoretical, topical, and methodological trends of agenda-setting research over time from 1972 through 2015 with a thematic meta-analysis. They found that traditional agenda-setting theories such as first and second-level agenda setting, inter-media agenda setting are still important theoretical framework in empirical research. Meanwhile, the authors also noticed that agenda setting research has extended from mainly focusing on politics to a more diverse range of research topics, including environment, health, and social issues. Yet content analysis has been the dominant method used by agenda setting scholars, which also indicates the importance of methodological advancement in the field of Natural Language Processing. However, missing from the empirical research on agenda setting is a more dynamic perspective of measuring the temporal features of the agenda setting process. Luo et al. (2019) conducted a meta-analysis on empirical studies published during the same period, between 1972 and 2015. The review included 67 studies that found the strength of relationship between public and news media agendas has been consistent, with mean effect size equals to 0.51 (see Wanta and Ghanem 2007, for a meta-analysis on 90 agenda-setting research, which suggested an overall mean correlation at 0.53).

Previous literature reviews of agenda-setting research also examined different types or levels of agenda-setting effects. For example, one review article focused on the new theoretical framework of inter-media agenda setting (IAS) and reviewed IAS research published from 1997 to 2019 (Su and Xiao 2021). The review showed that IAS research emphasised the relationship between traditional media outlets, with the most common agenda flows were studied between two traditional media. The exchange of agendas between a pair of emerging media (digital platforms such as social media) is the least commonly observed by researchers. Moreover, methods such as Granger-causality or ARIMA are recommended by the authors to supplement traditional methods of time serious analysis such as cross-lagged correlation. Markiewitz and Arendt (2020) provides a

review on the pioneering work in Agenda Setting research and newer avenues of research and developments on a conceptual level, such as second-level/attribute agenda-setting, agenda-building and agenda-melting. The authors also rightly pointed out that the wide spread of social media means a revision of agenda setting theory, in particular, testing or theorizing the relationship between traditional and new media in setting agendas for the public.

Nevertheless, there is a gap in literature reviews about agenda-setting empirical research: While an increasing number of empirical research studies have applied large-scale data sets and computational methods in examining agenda-setting effects, very few review articles systematically evaluated empirical research in this field. We aim to fill the gap in literature: In our systematic review, we aim to understand how the theory of agenda setting has been rationalized and revised in empirical research that is based on large-scale data sets or computational methods. Moreover, we are interested in summarizing important methodological gaps or advancement when using big data to understand media and public agendas.

Research questions

Our goal for this review is not to replicate or update literature reviews mentioned in the previous section, but rather to take a targeted look at the impact computational methods and large-scale data are having. We aim to provide social scientists and computer scientists alike with an understanding of what methods and data sources are commonly used, identify limitations of those methods and data sources, and encourage new, innovative cross-disciplinary research. This review paper examines agenda setting literature from 2016 to 2021 that employs large-scale data and/or computational methods. Our overarching research question is to understand: How is the "media agenda" identified, measured, and compared in the era of large-scale data and computational methods. We address this question through examining three sub-questions:

RQ1 What is the subject of analysis in agenda setting research, using large-scale datasets and computational methods?

RQ2 How is the media agenda measured and compared in large-scale empirical research?

RQ3 How do computational approaches inform the study of media agendas at scale?

Methods

Data collection

Data collection for this review includes both journal articles and conference proceedings. For journal articles, we used Scopus, Elsevier's abstract and citation database, to collect empirical studies that focused on media agenda-setting and applied quantitative or computational methods. We searched for two areas of empirical papers using Scopus and limited time frame to 2016–2021: The first area focuses on media agenda setting in the context of digital media platforms, using keywords such as "agenda setting", "agenda-setting", or "media agenda", combined with keywords such as "digital", "internet", "social media", "online" or "network", to search within titles, abstracts, and keywords of publications. The second area focuses on quantitative or computational approaches applied in agenda setting empirical papers, consisting of agenda setting keywords and keywords such as "comput*", "big data", "large-scale", "large scale", or "quantitative". We only included journal articles, conference proceedings, chapters, and books and removed research articles published in the research fields as defined by Scopus of medicine, health professional, nursing, immunology and microbiology, neuroscience, pharmacology, toxicology and pharmaceutics, and veterinary as their research focus is not on media agendas. There are 901 unique publications that match the two literature search queries on Scopus. We chose IEEE and ACM conference proceeding databases to search for empirical articles focusing on media agenda or agenda setting theory using quantitative methods. There are nine articles published on IEEE proceedings and six articles published on ACM conference proceedings with "agenda setting" or "media agenda" in their titles. One conference paper published in ICWSM proceedings, under the AAAI conferences, focused on agenda setting theory. The Scopus query used in the literature review is: (TITLE-ABS-KEY ("agenda setting" OR "agenda-setting" OR "media agenda*") AND TITLE-ABS-KEY (digital OR internet OR "social media" OR online OR network) AND NOT TITLE-ABS-KEY ("research agenda") AND PUBYEAR ¿ 2000) OR (TITLE-ABS-KEY ("agenda setting" OR "agenda-setting" OR "media agenda*") AND TITLE-ABS-KEY (comput* OR "big data" OR "large-scale" OR "large scale" OR quantitative*) AND NOT TITLE-ABS-KEY ("research agenda") AND PUBYEAR ¿ 2000) AND (LIMIT-TO (DOCTYPE , "ar") OR LIMIT-TO (DOCTYPE , "cp") OR LIMIT-TO (DOCTYPE , "ch") OR LIMIT-TO (DOCTYPE , "bk")) AND (EXCLUDE (SUBJAREA , "MEDI") OR EXCLUDE (SUBJAREA , "HEAL") OR EXCLUDE (SUBJAREA , "BIOC") OR EXCLUDE (SUBJAREA , "NURS") OR EXCLUDE (SUBJAREA , "NEUR") OR EXCLUDE (SUBJAREA , "PHAR") OR EXCLUDE (SUBJAREA , "VETE"))) Our first literature collection was conduction in May 2021, and the second round of literature collection was conducted in Feb 2022. An additional 33 articles that fit our search query were found in the second round of data collection and added in the review sample.
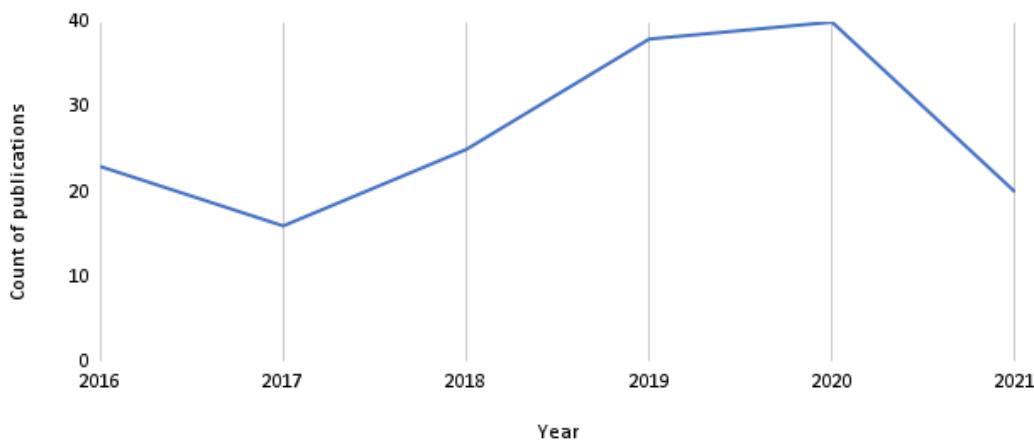
Coding scheme for the screening process

We started the systematic review process by developing a coding scheme to identify publications that either used a computational approach or analysed large-scale data in agenda setting research. Table 1 shows each coding scheme, descriptions, and whether or not the column is free-text in the first round of screening process. Two independent coders applied the coding scheme on empirical paper collected in the literature dataset. Intercoder reliability (ICR), measured by percent agreement is 99.9% on all the non-free-text columns in the coding scheme. After the screening process, there are 163 papers in total included in the final sample for the literature review.

Publications focused on media agenda setting theory and computational methods (or large-scale datasets) have increased since 2017 (See Figure 1). Eight out of 10 top-ranked journals or conference proceedings that published the highest number of media agenda setting and big data articles are communication journals. For example the top three journals that published most agenda setting and big data research papers are: Journalism Studies, International Journal of Communication, and Journalism and Mass Communication Quarterly. Inter-disciplinary conferences such as ICWSM (International AAAI Conference on Web and Social Media) and Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) are two proceedings in the top ranked journals/proceedings of media agenda and big data research.

Coding scheme for analyzing agenda setting papers

After identifying 163 publications that focused on agenda setting theory and applied computational methods (or used a large-scale data set in the research), we systematically coded all publications based on coding scheme shown in Table 2. We took an iterative procedure that involves two rounds of coding and re-coding. In the first round, five

Figure 1. Number of publications per year.

Table 1. Summary of the coding scheme for screening publications that either used computational approach or analysed large-scale dataset.

Coding category	Description	Levels
Relevant	The publication has to focused on agenda-setting theory AND (The publication used computational methods for data collection/analysis OR the publication used a large-scale dataset	Yes/No/Maybe
Data collected from news media or social media websites?	Whether or not data collected in the publication is from news media (newspapers, news websites) or social media platforms	Yes/No
Focused on media agenda-setting theory?	Whether or not the publication used agenda setting theory framework	Yes/No
Used computational methods for data collection or data analytic?	Whether or not the publication used computational methods to collect or analyze the data	Yes/No
Used large-scale data set (N more than 1000)?	Whether or not the data set used in the publication has a sample size equal to or greater than 1000	Yes/No

researchers divided the publications and coded the articles independently in both short labels and free-text, based on the coding scheme. In the second round of coding, we re-coded all publications into a smaller set of short labels by merging similar short-labels identified in the first round of coding. The benefit of using this two round, bottom-up coding system instead of a predetermined list of coding scheme is that all short labels used in the second round of coding are determined based on empirical studies.

Results

Keywords in agenda setting publications across six years

To understand important themes examined in agenda setting and big data empirical studies, we used Natural Language Processing (NLP) approaches to extract keywords from the abstracts of selected publications in this review.

Publications in January and February of 2022 were not included in computational linguistic analysis. We built the corpus for text analysis using abstracts of the publications on agenda setting and big data between 2016 and 2021. After tokenising sentences, removing common English stopwords, lemmatising words, identifying bi-grams (commonly appearing word combinations), and tagging Part-of-Speech and keeping nouns, adjectives, and adjectives only, we used the cleaned corpus for the final text analysis. We used TF-IDF, which measures the importance of words but controlled for their appearances across all documents, to identify top-ranked 100 keywords across all six years (See Figure 3).

Trends in agenda setting publications across six years

The visualization of the top-ranked keywords (See Figure 3), measured by TF-IDF scores, shows that agenda setting and

Table 2. Summary of the coding scheme for the systematic literature review.

Coding category	Description
Theoretical framing	Theoretical frame used in the study
Agenda setter	Who sets the media agenda
AS target/impact	The audience of media agenda
Time period: Length	Length of data collection time period
Platform(s)/Data basis	The platform/data basis of the dataset
Concrete data source	The specific data source of the study
Units of analysis	The unit of data analysis
Size of sample	The size of dataset
Country/region	Country or region of the study
Language of the source material	Language of data sources
Comparative?	Is the study comparative?
Methods for comparing agendas	How media agendas are compared
Qual? Quant? Mixed?	Is the study qualitative? quantitative? or used mixed methods?
Data collection strategy	How dataset in the study is collected
Specific issue vs full agenda	Whether the study focused on a specific issue or full agenda of media
Method for identifying issues/agenda	How issues or agendas are identified
Categorization scheme	How media agendas are categorised
Method for identifying second level agenda	How second level media agenda is identified
Method for identifying third level	How third level media agenda is identified

big data research continued the previous research interest on "medium" and "news," which are ranked as the first and second most important keywords measured by TF-IDF scores. Both emerging media platforms such as "Twitter" and traditional news outlets such as "newspapers" are important arena of agenda setting research in the big data era. Meanwhile, we found that agenda setting studies, using large-scale data sets, have focused extensively on political realm, indicated by top-ranked keywords such as "political," or "party."

We are also interested in the increased or decreased importance of keywords in the abstracts of agenda setting and big data research across the six years. To explore this question, we calculated the percentage of agenda setting publications that contain the top-ranked 20 keywords in each year, but mainly compared the top 20 keywords in 2016 and 2021. Note that we only include keywords that appear in both 2016 and 2021 in the analysis. Figure 2 shows the changes of importance of keywords between 2016 and 2021. Some words, including for example, "politics," "newspaper," and "twitter" are on decline, while other topics, including for instance, "crises," "public," and "social" are increasing in agenda setting and big data research. Using qualitatively coding of all agenda setting and big data research across the six years, we will explore in-depth the trends of agenda setting research in the next section.

RQ1: What is the data sources of agenda setting research? What is the subject of analysis? Who are agenda setters and audiences of media agendas? What is the time length of data collection?
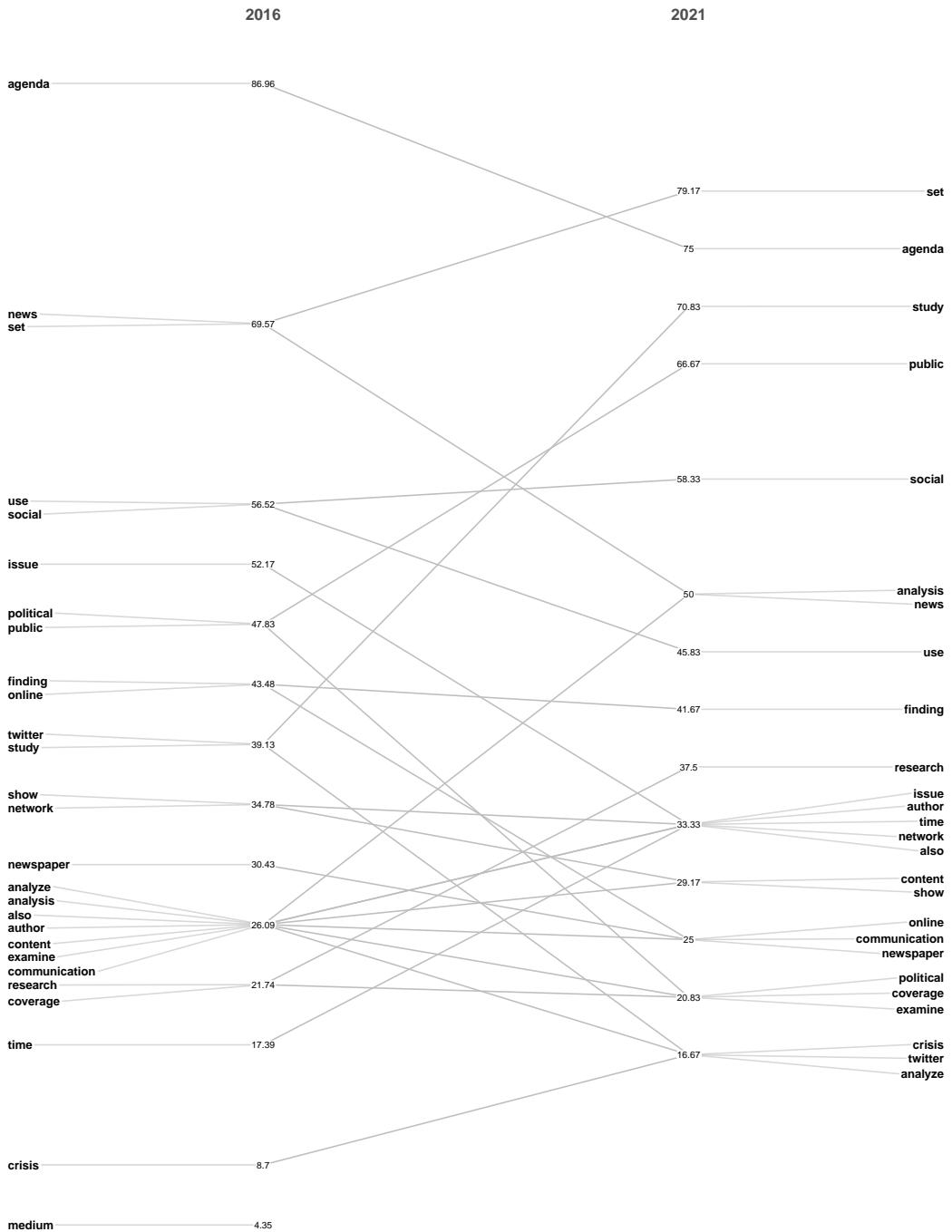
Our first research question asks the main subject of research in agenda-setting studies that used large-scale data sets. We approached this research question from data sources, levels of agenda-setting, specific vs full issue agenda(s), main agenda setters, and main targets of media agendas. By systematically reviewing and coding these dimensions, we are able to explore **which** media outlet or digital platform(s) is commonly used to study media agendas at large scale, **what** media agendas have been studied and **who** have been identified as agenda setters or targets of agenda setting.

Data sources: Traditional media vs social media. Empirical studies on agenda-setting collected research data from two main sources: traditional news media such as newspaper and social media platforms such as Twitter. We found that over half of reviewed studies relied on one type of data, with the remaining studies based on triangulation of traditional and social media data sets.

In line with the historical origins of agenda-setting theory that analyzed traditional media, 89 out of 163 (54.6%) of the reviewed studies included either online or printed versions

Change of the percentage of top 20 keywords appearing in agenda setting & big data publications in 2016 a
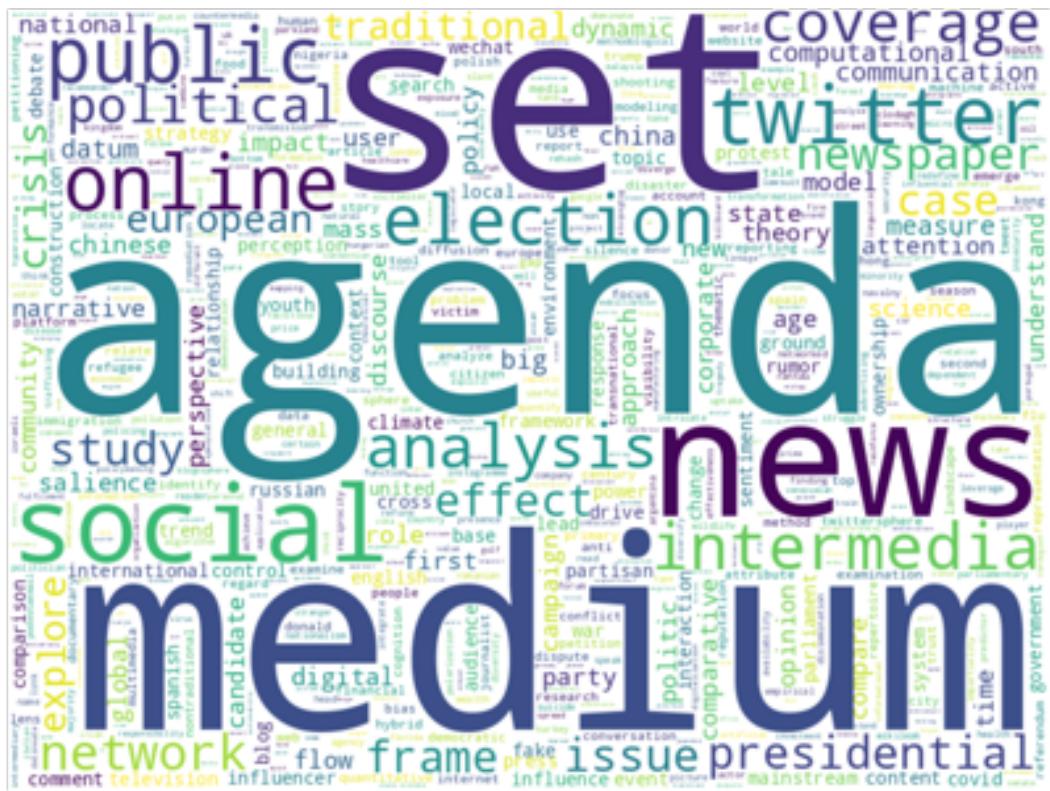
Top–ranked keywords measured by TF–IDF scores and only keywords that appear in all six years are shown

2016	2021

agenda — 86.96

79.17 — **set**

75 — **agenda**

70.83 — **study**

66.67 — **public**

news
set — 69.57

58.33 — **social**

use
social — 56.52

issue — 52.17

50 — **analysis**
news

political
public — 47.83

45.83 — **use**

finding
online — 43.48

41.67 — **finding**

twitter
study — 39.13

37.5 — **research**

show
network — 34.78

33.33 —
issue
author
time
network
also

newspaper — 30.43

29.17 — **content**
show

analyze
analysis
also
author
content
examine
communication
research — 26.09

25 — **online**
communication
newspaper

coverage — 21.74

20.83 — **political**
coverage
examine

time — 17.39

16.67 — **crisis**
twitter
analyze

crisis — 8.7

medium — 4.35

Figure 2. Slope chart of top 20 keywords between 2016 and 2021.

of newspaper. Hence, the most common units of analysis are content of news articles, which was featured in 63.8% of sampled studies. Other units of analysis included five articles based on comments on news articles, for example Maniou and Bantimaroudis (2021), Rosino and Hughey (2017), and Koltsova and Nagornyy (2019). We found that empirical articles that focused on traditional news media frequently relied on newspaper data bases as their data sources, with 50 out of 89 traditional news studies collected data by querying a newspaper database. The most popular databases were LexisNexis and Factiva which were used in 19 and 8 empirical studies, respectively. Other, less frequently used databases included some with content similar to LexisNexis and Factiva, such as NewsBank, ProQuest, and Access World News, but also some focusing on non-English sources, such as Swissdox and WiseNews. In only two cases did the researchers use a second-hand data set and only one study mentioned outsourcing the news data collection entirely.

Figure 3. Word Clouds of Keywords, Weights represent TF-IDF scores.

Following news media, social media are the second most used source of data. They were used in 57 out of 163 studies, of which more than half overlapped with studies that looked at traditional news media data. For social media data, the individual post (or tweet) are the most common units of analysis. Unlike news data that was often collected via databases, social media data was almost always directly sourced from the targeted platform. Social media's own APIs (for example, Twitter, VKontakte) and social media scraper (for example, WeChat and Weibo) are common methods of data collection from social media platforms.

Meanwhile, the preferred platform for social media data by a large margin is Twitter, as the micro-blogging platform provided data for 30 empirical studies.* Similarly, Weibo (a Twitter-like social media platform) is also used more frequently used for collection research data than other mobile social media applications such as WeChat. It is worth noting that Twitter went from a relatively minor category in 2017 to the most-used data source in 2020. Runner-up platforms are remarkably less popular, with Facebook and Weibo featuring in five of studies respectively. Other social media platforms used were Instagram (Towner and Muñoz 2020), Quora (Shi et al. 2019), VKontakte (Judina and Platonov 2018) and WeChat (Cong et al. 2019) but none of them were used in more than two studies. There is a lack of empirical studies on agenda setting that mine social data from digital platforms other than Twitter.

Eight studies analyzed search data, including tool kits such as Baidu Index, Google Trends and Google Advance search. Search engines such as Google and Baidu provide "long-term evidence of the keywords searched by users around the world" (Maniou and Bantimaroudis 2021, p. 1131) and offer an alternative method of measuring public agenda at scale (Sun and Zhong 2020; Pinto et al. 2019). Nevertheless, using large-scale search data also has some limitations: search indexes are often shared by search engine platforms in an aggregated format, and thus causal inferences about the "individual-level psychological processes cannot be made." (Lee et al. 2016, p.451).

Compared to early work on agenda setting that focused on the comparisons of media agendas measured in newspaper content and public agendas measured by surveys, we found survey methods were used in only ten studies and only four studies used first-hand survey (Wu and Guo 2020; Jonkman et al. 2020; Apuke and Omar 2022; Cheng 2016). Also, survey data set were all used in triangulation with large-scale data set. This shift away from surveying, a historically

---

*The studies we analysed predate actions Twitter took in spring 2023 to restrict API access. Those restrictions will likely impact the prevalence of Twitter in future research.

essential as a main datatype to agenda setting research, emphasizes the growing importance of online platforms as data collection sources for agenda setting research. Other historically important sources, such as television and radio also seemed to garner little interest. Only three studies looked at television- or radio related material (Harder et al. 2017; Weimann-Saks et al. 2016; Rogstad 2016). Political information sources were also under-studied: We only identified few studies that used political data from outside social and news media: two studies mentioned using political press releases (Gilardi et al. 2022; Jansen et al. 2019), two used parliamentary data (Males and Van Aelst 2021; Herrero-Jiménez et al. 2018). A review of theoretical debates around media and public agendas suggested that future empirical research can explore the triangulation of large-scale data with public opinion polls or data set collected directly from political parties (i.e., press release, party monographs etc).

Research subjects: Locations and languages of media agenda research The locations and geographic scopes analysed in our sample varied vastly in scale, ranging from studies that focused on sub-regions of a country (e.g., the province of Aydın Turkey in Yüksel and Dingin 2021) to studies that covered multiple countries (e.g., Sintes-Olivella et al. 2022) or took on an international scope (e.g., Kawakami et al. 2020).

The scope of location used for analysis was often relative to the issue under study. If the issue concerned only a country or a sub-region of one country the researchers typically disregarded data beyond that region, while studies that analyzed issues that concerned multiple countries or the whole world (e.g., migrants or natural disasters) were more likely to include groups of countries or forego geographic limitations to their data collection process altogether.

Among the countries under study, the USA had the most frequency of appearing in our sample with 28 out of 163 empirical research focused on media agenda in the US, followed by China (18 out of 163), and the Netherlands (8 empirical studies). Research interest in China has been growing from 2 studies in 2016 to 7 in 2019 but dropped to 5 in 2020. Research focus on the US remains the highest across all countries, with more than 4 studies on average published on the media agenda in the US every year. Meanwhile, we also observed an obvious lack of attention on so-called 'Global South' countries such as India, Indonesia, and Pakistan, all of which received 0 academic attention in agenda setting and big data research. We therefore suggest future agenda setting scholars to also include developing

countries and regions in empirical research of media agendas at large scale.

Most studies focused on a single country (73.6%), a small number of studies focused on two or three countries (6.75%) and another handful (6.70%) focused on more than three countries. Studies that looked at traditional news data showed a slightly larger variety of locations compared to studies that looked at social media data. For example, Jiang et al. (2021) studied a international conflict and how media influences public agendas on the conflict across China, the US, and South Korea.

When it comes to the study of multiple country studies, our sample included many comparisons between regions that are geographically close (e.g., US/Canada, Germany/Austria, and Russia/Ukraine). Moreover, only 7 studies explicitly mentioned looking at countries from at least two different continents. More than 26 sampled studies showed no clear geographic orientation. To further examine the geographic focus of agenda setting research, we examined languages studied in the sampled studies.

We identified 24 different languages used in our sample. English was found to be the most common (67 out of 163 studies) followed by Chinese (19 studies). Other notable languages were Spanish (9 studies) and Dutch (8 studies). Most studies tend to focus on source material in a single language, with only 17 out of 163 studies comparing sources across two different languages. Seven studies looked at Chinese and English source material simultaneously, making the English-Chinese cross linguistic comparative studies to be the most commonly observed multi-lingual agenda setting research. The over-dominance of English as research data could be partly explained by the geographical focus on the US in agenda setting research. Meanwhile, data sources such as international platforms (Twitter) were also frequently used to study media agendas in countries where English is not the official language.

Agenda setters and audiences of media agendas Original theoretical framework of agenda setting was applied to traditional media outlets such as newspapers, TV and radio stations, when both media professionals (journalists or news editors) and government officials could be agenda setters. We examined who are agenda setters in the selected empirical studies and found that the importance traditional media as agenda setters persists. In our sample 68 out of 163 studies identified news media as the main agenda setters in empirical research. In comparison, social media were only indicated to be the main agenda setters in 13 studies. In a further 14 studies, social media and news media were both

defined as agenda setters. Public participation in social media and content generated by users have been theorised as a phenomenon known as reverse agenda setting. We found 15 empirical research in our sample listed media audience or the public as agenda setters. Interestingly, although early studies of agenda setting process highlighted the importance of parties or politicians in setting media agendas for the public, we found that there has been a decline in studies of politicians as agenda setters (from 5 studies in 2016 to less than 2 studies in the following years).

Whereas the agenda setter was identifiable in all sampled studies, the targeted audience was less so. In 35 out of 163 of studies we found no clearly defined target. In the remaining studies, the main targets were news media and its consumers (35 studies), as well as the public (40 studies), and social media users (15 studies). Government and political actors were mentioned as targets in 7 out of 163 studies. The increasing fluidity of agenda setter and target of media agendas indicating that empirical research focusing on media agenda in the digital era need to widen the definition of agenda setters and agenda audiences and theorise the flexible boundaries between media setters and audiences.

Time length of data collection Using computational techniques, agenda setting scholars are capable of analysing data set from the longitudinal perspective. We identified 18 publications in our reviewed empirical research collected data for 365 days (one year), which is the most commonly used window of data collection period. The second most commonly used data collection period is 30 days (one month), and this data collection length was used by 8 publications. Even though the publication we included applied agenda setting research on large-scale data, less than one third of sampled studies collected data set for longer than a year (50 out of 163). The longest time period collected in the sampled studies is an empirical research that studied ten television documentaries, with the oldest traced back to the 1970s (Omoya et al. 2020). The second longest time period collected is an empirical study that examined corporate scandals between 1990 and 2016 (Barkemeyer et al. 2020). The third longest time period is 15 years, where scholars compared the media agendas of climate change in Irish newspapers between 1997 and 2012 (Wagner and Payne 2017). We believed that with the assistance of computational techniques such as time series analysis, data visualisation, and computational linguistics, such longitudinal study of media agenda setting could contribute to the understanding of the dynamics between the media and the public across time; yet, there is currently a lack of agenda setting research focusing on long period of time.
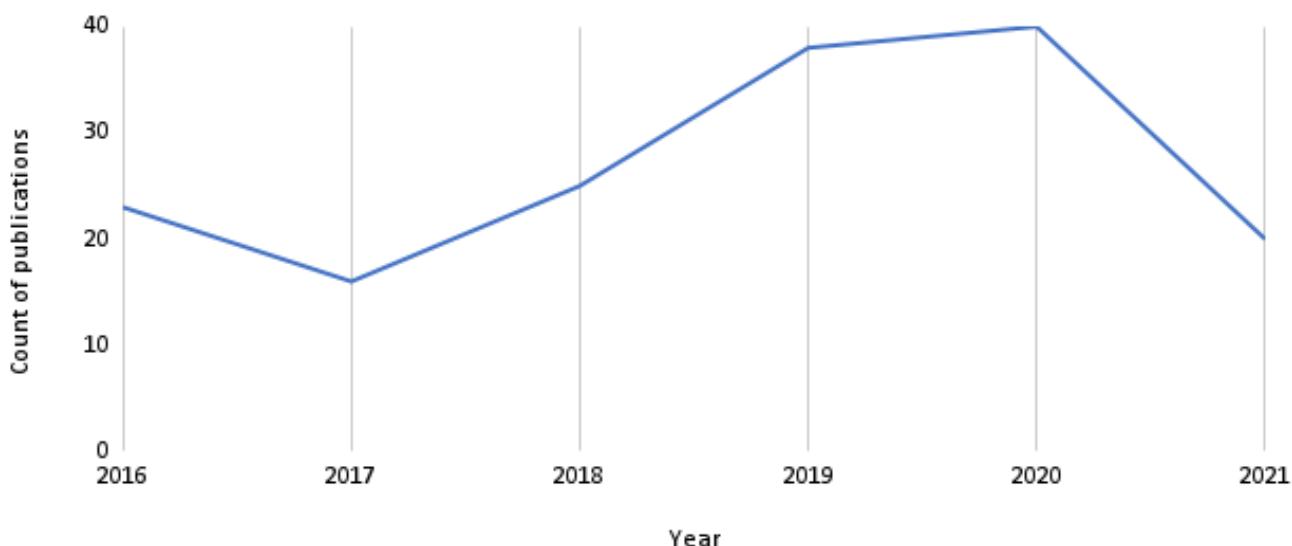
RQ2: How is the media agenda identified, measured and compared in the era of large-scale data and computational methods?

Measuring media agendas: Choosing between qualitative or quantitative approach? In agenda-setting research, both quantitative and qualitative analyses are widely used approaches. Of the 163 sampled studies, the majority of (103 publications) employed quantitative methods, with only ten studies opting for purely qualitative methods and 39 using a mixture of both approaches. This skew towards purely quantitative research was stronger in studies that looked at social media data, where 76.6% of studies were purely quantitative research. For studies that looked at traditional news media data, this percentage was 64.4%. It should be noted here that the quantitative skew is not representative of agenda setting as a field, as it is (partially) the result of our selection criteria, which required the use of large-scale data and/or computational methods. Qualitative approach helps add interpretations of the nuances of media agendas in empirical research. For example, Dekker and Scholten (2017) applied qualitative comparative analysis on media coverage of sixteen events related to immigration and revealed media dynamics through in-depth reading of content. Thus, we recommend that advanced computational approaches such as unsupervised machine learning of media agendas not replace close reading of media content.

Single issue or full issues of media agendas When it comes to the focus of agenda setting research, an important decision made is whether or not the researchers focus on a clear specific issue. Issues we encountered ranged from one or multiple persons (e.g., politicians), events (e.g., Chinese National Day), organisations (e.g., companies), a group of users (e.g., young people), or a specific type of information (e.g., online rumours). The majority of sampled studies (more than 112 papers) focused on analysing a single issue of agenda setting rather than describing the full agendas on the media outlets. The lack of comparing across diverse issues may lead to poor theoretical generalisability of empirical findings. While the targeting benefits of issue selection for focused research are clear, a researcher might also opt for a more generative approach. Such a "full-agenda" approach sees topics emerge as a result of the research, which leaves room for unexpected insights.

Both the studies that used news media data and those that used social media data adhered to a similar distribution

Figure 4. Number of publications across different data collection period lengths

of specific issue and full-agenda approaches, with studies looking at social media data only slightly more skewed towards full-agenda research. This might be explained by the relatively easier data access provided by social media platforms that enabled researchers to track and analyse more than one media agenda issues. A bigger difference is seen when compared against the three levels of agenda setting analyses. More than half of the full issues research is done from a first level perspective. Full-agenda type research was especially uncommon in second level analyses. The difficulty of diving deeper into the content or network analysis of second- and third-level research in full issues research is partly due to restrictions in manuscript length, which does not allow rich content analysis of each media agenda identified in full issues research. However, limiting full issue research to first-level analysis also means a lack of thorough examinations into the content nuances or the structural dynamics between different media agendas.

Finally, the biggest methodological difference between specific vs full issue media agenda research lies in their methods for data collection. Research focusing on specific issues almost always uses some form of keyword query. Alternatively, the majority of full issue research (over 50 per cent of full issue research) centers its data collection around gathering all output of a specific news source, social media account or company for a given period of time. For example, Bright (2016) collected a large sample of news articles from the front page of BBC News by crawling the BBC front page once every 30 minutes during a two week period. The resulting list of news agendas include a diverse range of news

topics, which enables a full issue research of media agendas. Such a comprehensive approach of collecting full issues was rarely used in studies focusing on a specific issue.

Theorising and measuring three levels of agenda setting
Previous agenda setting research has highlighted three levels of agenda settings: first-, second- and third-level analysis of media agendas. In the reviewed empirical research, we find the majority (116 papers) of empirical studies focused on one level of agenda setting, the remaining studies (31 papers) included some mixture of two or all three levels of agenda setting. The lack of multi-level empirical studies suggests that future research project could focus on comparing various levels of agenda settings. First-level agenda setting is the most researched levels of agenda setting empirical research, with 80 empirical studies focused on this level. Meanwhile, beyond the well-established three levels of agendas, we also find some empirical studies that have examined media agenda via inter-media agenda setting (Conway-Silva et al. 2018) or reverse agenda setting perspectives (Luo and Harrison 2019). When compared over time, third level analyses seem to have peaked in popularity around 2019. Second level analyses rose in popularity from 2016 to 2020, but dropped in 2021. Finally, first level analyses recovered from a near 50 percent drop from 2016 to 2017, but also dropped in 2021.

The most common method of extracting first-level agenda issues remains hand coding, though its popularity has been declining. This method was employed in 41 studies that analysed first-level agenda setting. This method is followed by simple frequency calculations, based on keywords or

string searches, which was used in 26 articles examining first-level agenda setting. Topic modelling for first-level analyses took off starting in 2018: 17 of studies that included first-level analysis employed topic modelling methods, of which more than half used Latent Dirichlet Allocation (LDA) topic modelling. Finally, website or social media meta metrics were used in 16 of first-level analyses.

Similar to first-level agenda setting research, the most common method of analysing the second level of agenda setting remains hand coding (used in 25 empirical studies), though there is less of a decline over time. However, unlike first level media agenda analyses, keyword/string frequency analyses are much less common and were done in only two studies of second level analyses. Computational methods such as sentiment analysis are featured in 14 studies. Finally, topic modelling is used in 8 second-level studies. Compared to first-level empirical research, studies that explored second-level media agenda showed a higher level of research interest in both topic modelling and sentiment analysis.

We also compared methods used to analyse the third level of media agendas. Network analysis and quadratic assignment procedure (QAP) methods were applied in analysing third-level agendas. We therefore observed a differences between content- and network-driven research approaches between second- and third-level of media agenda setting research. Unlike for the first two levels, studies that employed third-level agenda analyses were much more computationally driven. More specifically, 20 studies that analysed third-level agenda setting used some forms of network analysis such as issue or attribute co-occurrence networks or eigenvector centrality analyses. Hand coding was used in only six third-level agenda setting empirical studies.

Finally, the more direct methods that are historically associated with agenda setting research, such as surveying and interviewing, were scarcely employed. Surveys were used twice for both first- and third-level analyses and only once for a second-level analysis. Interviews were conducted only twice in total: once for a first level and once for a second level analysis. The lack of interviewing makes sense given our inclusion criteria, but we did include surveys with at least 1,000 respondents. Meanwhile, we also noticed a lack of computational methods applied in large-scale data analytic of agenda setting research, with hand coding used as the most common method of identifying first-level agendas and analysing second-level agendas. However, we also propose that future empirical research on agenda setting could combine the rigid process of hand-coding of agenda data with data-driven approaches such as

supervised or unsupervised machine learning methods, to provide validation process following computational analysis of media agendas.

Comparative methods Early empirical research and theorisation of media agenda setting theory emphasises the competition of public and media agendas in the marketplace of attention, and therefore highlights a comparative perspective behind agenda setting research. Recent decades have witnessed the competition of public attention between digital and traditional media outlets, which has in turn lead to comparative studies of new vs traditional media agencies. Reviewing how different media agendas are compared in empirical research and what methods are employed to analyse the similarity and differences between media agendas is an important step towards understanding the use of big data in agenda setting research.

In our 163 reviewed papers on agenda setting research, 159 articles compared agendas. Among them, the most popular approach used in empirical research is comparing multiple information sources or agenda setters (over 60 per cent). Meanwhile, comparisons across time (longitudinal) and across spaces (geographical) are two other common perspectives of comparing media agendas, with each type of comparative study accounts for over 20 per cent and 13 per cent respectively.

When it comes to specific comparison methods, the most frequently used approach is statistical analysis, accounting for 67.9% of our sampled studies. However, among the articles that used statistical analysis for comparing media agendas, descriptive analysis is the most frequently applied analytical approach (50.0%), followed by followed by correlation analysis (25.0%) using Spearman's correlation or Pearson's r correlation. Models such as Vector autoregressive (VAR) model, longitudinal model, panel model, and diverse regression models account for 23.1% of total empirical studies, followed by statistical tests such as ANOVA test, Chi-square test, Portmanteau (or Q) test, which accounts for 15.7% of studies.

The second most commonly used comparative method is time series analysis (48.1%), which is mainly used by researchers to track the dynamic temporal flows between different media outlets at various time points. Among studies that applied time series analysis, Granger causality (42.3%) was the most commonly used method, with five empirical studies using cross-lagged correlation and four cases employed the Dick–Fuller test. Other more advanced methods also seen for time series analysis gradually included but are not limited to ARMA model, ARIMA model, and

time lag methods, providing models for making predictions or causal inferences.

In addition to statistical analysis and time series analysis, over a quarter (26.4%) of empirical studies applied social network analysis (SNA) for comparative study of media agendas. The most frequently applied method is Quadratic Assignment Procedure (QAP), which accounts for 33.3 per cent of studies that adopted social network analysis method. Multiple Regression Quadratic Assignment Procedure (MRQAP) and co-occurrence network analysis as the second and third commonly used methods, each representing 11.9% of empirical research in our sample. Other methods used in social network analysis of media agendas include measurement of media network centrality (eigenvector centrality) or semantic network analysis of media content. Meanwhile, network visualisation provides a powerful toolkit to represent differences and similarities in the structures and content of media outlets. Such a visualisation of comparative analysis of media agendas can be analysed in open source software, such as the OpenOrd algorithm, the HITS algorithm, the NetDraw software and the R package "igraph".

Comparing to quantitative methods used in comparative studies of media agendas, very few empirical research applied qualitative methods in comparing media agendas. Nevertheless, most empirical research that applied qualitative comparisons of media content did not document details of the qualitative method designs. Also missing from empirical research of media agenda comparisons are methods such as machine learning or spatial (geographical) analysis. We believe that computational linguistic methods such as Natural Language Processing has potentials in cross-media, cross-country, cross-time comparisons of media agendas. Future empirical research on media agendas can explore the use of computational linguistic methods in addition to statistical, time series, or social network analysis of media agendas.

RQ3: How do computational approaches inform the study of media agendas at scale?

Unsupervised models in identifying and analysing media agendas and validation process Computational approaches such as social network analysis, natural language processing, and computer visions are increasingly applied in the field of information and communication research (Hofman et al. 2021). Agenda setting scholars have applied automated text analysis and unsupervised methods to identify media agendas, but as Grimmer and Stewart (2013)

point out, empirical research often lacks a validation process when using automated text analysis. We closely reviewed the 19 articles in our data set that used automated text analysis (i.e., unsupervised models, topic models, or LDA) and found only 5 out of 19 articles contain a validation process after applying automated text analysis. Manual validation is the most commonly used approach for checking topic validity. For example, Aslett et al. (2022) inspected tweets that fall into each topic under a relatively low probability threshold and to check if their topic model accurately categorised media agendas. Manual validation is also used to collapse topics that share similar themes and filter out topics that lacks human interpretable themes (Bozarth et al. 2020; Su et al. 2020; Wang et al. 2021; Dai et al. 2021). Nevertheless, other validation approaches, such as validating identified topics via external events (comparing attention to media agendas vs. timeline of major events during the same research period) has not been adopted by media agenda scholars to date.

Categorisation schemes As mentioned in the previous section, many agenda setting studies rely on categorisation schemes (or "code books") to classify media agenda text into different topics. We found that more than half (57.1%) of the sampled studies used some form of a categorisation scheme, code books or references to label text into different categories of social issues, with 42.9% used their own original scheme to classify issues. In studies where authors used their own categorisation scheme, classification based on lists of keywords was common. Among them, some authors also developed secondary categorisation schemes based on existing empirical research or standards. We also found that the use of secondary categorisation schemes was more common in second-level media agenda setting research than for first- or third-level agenda setting research. Few empirical research combined categorisation schemes published in previous studies with categorisation schemes developed by the authors. Exceptions include Platonov and Svetlov (2021), who examined political communities on a Russian social networking website using addictive regularisation algorithm.

In these studies the researchers often came up with a list of categories based on the topic of the study and the research questions. In total, 14.1% of the studies used a secondary scheme to classify their sample content. These schemes were either inspired by previous research or standards. Though secondary schemes are relatively more used for second-level analyses than for first- of third-level analyses, original scheme usage was still far more common. Three studies mentioned using a combination of original and secondary schemes to classify their data. These studies usually used

categories established by previous agenda setting studies to develop a categorisation system of their own.

Only a handful of the studies that introduced an original categorisation scheme established frameworks that could be extended to future research. We argued that by comparing pre-existing categorisation scheme with coding schemes used in original empirical research, for example, based on supervised or unsupervised machine learning, agenda setting researchers can work collectively on enhancing the quality of categorisation schemes used to identify and cluster media agendas. Such a comparison between first- and second-hand categorisation schemes not only requires mixed methods design that involve both qualitative text analysis and computational data analysis, but also calls for an iterative research pipeline that consistently works on refining the training models or code books used by media agenda setting scholars.

Discussions

The digital revolution has created a hybrid media landscape that combines the 'older' and 'newer' media logics, which has spawned the need for more versatile methods to understand "agenda setting" theory. Our study delves into media agenda in the era of big data based on systematically reviewing previous empirical research. Our review indicated that since 2016, agenda-setting research has expanded in scope, topics, methods, as well as techniques.

Theoretically, agenda setting has become more complex and refined compared with the pre-big-data era. The initial agenda-setting studies focused on examining political elections. However, the emergence and rise of digital technologies has opened up new arenas for agenda setting research as information dissemination outlets are considerably enlarged by emerging media, social media in particular. On the one hand, research on traditional media agenda setting effects remains mainstream, while the discussion of intermediate agenda setting and reverse agenda setting has increased. The Internet has revolutionised the media landscape and spurred new questions about the interactions between traditional and emerging media. From the media users' perspectives, media content has shifted from a "pushed" flow of news feed to a "pulled" news environment where social media users are playing more important roles in tailoring the news content on their internet devices. Meanwhile, from the media ecology perspective, influential social media opinion leaders, digital-born news outlets, and independent journalists are also competing for the already-scarce resources of audience attention. Our review of big

data research in media agenda setting theory shows the rich and complex dynamics between traditional and social media in setting media agendas, while an increasing number of empirical studies confirmed a social-to-traditional media influence (Carazo-Barrantes 2021; Černỳ and Ocelík 2020) and sophisticated trans-media network of media agenda interactions (Guo and Vargo 2017; Funk and McCombs 2017; Vonbun et al. 2016). Computational methods such as social network analysis and statistical models such as time series analysis have contributed to the theorisation of such a "reversed" agenda setting process, providing new evidence to agenda setting mechanisms in the digital era.

On the other hand, we noticed a broader scope of research topics included in agenda-setting research using large-scale data sets. Powered with computational methods and tool kits, empirical researchers have extended the boundary of media agenda from political, economic, and social issues, to non-traditional media agendas such as rumours (Guo and Zhang 2020) or misinformation (Guo and Vargo 2020). Underlying the broader definition of media agenda is a shift of academic focus from traditional mainstream media outlets, to also include alternative media channels such as partisan media, conspiracy theorists, or bots on social media.

In addition to shifts of research focus on media agenda audiences and agenda setters, we observed a tendency of shifting from researching full media agendas to exploring specific issues or agendas on media. This trend seems to be contrary to the general increase of sample sizes in big data research of media agendas. However, it is partly due to the increasing research interests of communication scholars to focus on context-specific, country-specific, or domain-specific social issues. Meanwhile, it might also be explained by the fact that keyword lists or dictionaries are used to collect research samples from social media platforms,

Furthermore, while the development of social media has opened up the possibility of applying large-scale data, more and more research is shifting from full agendas to discussing specific issues. This may be due to the growing awareness among scholars that the effects of agenda setting are issue-oriented, subjective-differentiated or context-specific, i.e., require case-by-case consideration. At the same point, however, it's also technically easier to collect social media content with keywords than to collect the "full agenda." It is worth noting that, along with the evolution of agenda-setting theory, subsequent studies should be wary of narrowing the research issues and falling into repetitive studies.

What's more, over the past decade, the theoretical and geographical map of agenda setting has continuously evolved alongside changes in communications technology

and social trends. It has been pictured in more and more countries in North America, Asia, Europe, Africa and Latin America. However, it is undeniable that academic study of agenda setting is not evenly distributed globally. There is a notable lack of attention to countries in the so-called Global South, and more comparative approaches can be used in the future to achieve contextual diversity.

Methodologically, the Internet and social media platforms have expanded the methodological toolkit of agenda setting research, ranging from traditional content analysis and survey (McCombs and Shaw 1972) to big data analysis with computational methods on Twitter, Google, and so forth. For example, there is an increase in the application of machine learning methods (Su et al. 2020; Hemphill and Schöpke-Gonzalez 2020), the use of causal inferences in agenda comparison (Cao et al. 2021; Van Den Heijkant et al. 2019), and the application of network visualisation. In addition, many agenda-setting scholars have resorted to advanced time-series analysis and regression models. Nonetheless, many methods are simple keywords or off-the-shelf machine learning tools: these are not inherently problematic but there is a need to better evaluate the accuracy of these methods. Such evaluation may show opportunity for more advanced method development.

There are significant opportunities for further research in exploring data sources beyond the mainstream social media and traditional news sources. The current over-reliance on textual data, most notably those derived from Twitter, runs the risk of excluding various demographics across the globe. It is thus advised that future research not only shifts its focus to different social media but also takes the leap into multimodal (i.e., audio-visual) analysis to include the content of increasingly popular platforms such as TikTok, Instagram, or YouTube. We note that this review only includes English-language publications, and this creates a possible bias in the sampling process that could impact the results of our analysis. Future research could review agenda setting literature in non-English languages.

Conclusions and future directions

To conclude, the article explores the evolution of agenda-setting theory in the context of big data. We notice the expanded score of research in the field since 2016, particularly, researchers have broadened the focus of agenda-setting theory to new media environments (social media and search engine) and to emerging dynamics between public and media agenda-setting effects (reverse agenda-setting). The rise of digital platforms not only lead to new data sources

for empirical research but also necessitate a shift in theory building to explain the complex interactions between various media sources and audiences in digital sphere.

We observe great shift in research methodology such as computational techniques and large-scale data analytics. Nevertheless, the review also suggests a critical need for theoretical innovation, especially in integrating new media agents into the agenda-setting theoretical framework. This integration is important to enhance the relevance and inclusivity of the theory in the digital era. We argue emerging non-traditional media sources like digital-born or partisan media (e.g., Breitbart) and independent social media users warrant close scrutiny. Moving forward, the future framework of agenda-setting theory for the digital era should be tailored to accommodate these new players (agenda-setters) within its structured framework, ensuring a comprehensive representation and understanding of contemporary media dynamics.

The evolution of agenda-setting research in the digital era underscores a promising intersection where computer scientists and social scientists can collaboratively advance our comprehension of the attention economy. This interdisciplinary collaboration has the potential to combine the computational prowess of big data analytics with the nuanced understanding of human behaviour and societal dynamics.

However, merging these distinct paradigms is not without challenges. Computational tools, while proficient at identifying media-covered issues, often overlook the nuanced editorial processes that mould media agendas. Conversely, social scientists, when delving into large data sets for agenda-setting, sometimes remain tethered to traditional hand-coded analytics. This approach, while thorough, can limit the full potential of computational methods, which offer rapid processing and fresh insights. Future research on agenda setting in digital era can improve the development of models that incorporate not only linguistic features but also journalistic and editorial nuances in analyzing media agendas, offering a more comprehensive view of media agendas. Exploration of media agendas beyond mainstream social media platform, as well as addressing non-textual media agenda data such as images, podcasts, or videos and short-videos can provide rich insights into a wider range of attention dynamics.

In our globalised world, media systems vary across different countries and media systems. Some platforms, such as Twitter or YouTube, transcend national boundaries and have been used by agenda-setting scholars for understanding the global information flow and media agenda dynamics. However, Western-centric platforms provide only a snapshot

of this vast media ecosystem. Platforms like China's WeChat, Russia's VK, or Japan's Line yeild significant influence in their regions, and their inclusion in agenda-setting research can offer a more holistic view of global media dynamics. By delving into these global networks, we can explore how local contexts (heterogeneity) both shape and are influenced by overarching global trends (homogeneity). A diverse geographical perspective can shed light on regional media systems, global network of information dissemination, and the interplay of developed and developing attention economies. Yet Big data analytics could be a powerful research method in this endeavor, with its capacity to process vast multi-country, multi-platform and trans-media datasets from diverse information sources, it can map the complex interconnections of global communication networks, track the real-time ebb and flow of information across regions, and analyze content that spans multiple languages and cultures.

By harnessing the combined strengths of theory and empirical research, social science and computer science, traditional and new media, Global North and South, agenda setting research powered by big data offers great opportunity for information, communication and technology scholars. Methodological and theoretical innovations in the field will reveal the profound impact of digital media on human behaviours and information society.

Acknowledgements

References

Apuke OD and Omar B (2022) Media and conflict reporting: a content analysis and victims assessment of media coverage of the conflict between farmers and herdsmen in nigeria. *Security Journal* 35(2): 345–366.

Aslett K, Webb Williams N, Casas A, Zuidema W and Wilkerson J (2022) What was the problem in parkland? using social media to measure the effectiveness of issue frames. *Policy Studies Journal* 50(1): 266–289.

Barkemeyer R, Faugère C, Gergaud O and Preuss L (2020) Media attention to large-scale corporate scandals: Hype and boredom in the age of social media. *Journal of Business Research* 109: 385–398.

Bozarth L, Saraf A and Budak C (2020) Higher ground? how groundtruth labeling impacts our understanding of fake news about the 2016 us presidential nominees. In: *Proceedings of the International AAAI Conference on Web and Social Media*, volume 14. pp. 48–59.

Bright J (2016) The social news gap: How news reading and news sharing diverge. *Journal of communication* 66(3): 343–365.

Cao RQ, Schniederjans DG and Gu VC (2021) Stakeholder sentiment in service supply chains: big data meets agenda-setting theory. *Service Business* 15(1): 151–175.

Carazo-Barrantes C (2021) Agenda-setting in a social media age: Exploring new methodological approaches. *The Agenda Setting Journal* 5(1): 31–55.

Černỳ O and Ocelík P (2020) Incumbents' strategies in media coverage: A case of the czech coal policy. *Politics and Governance* 8(2): 272–285.

Cheng Y (2016) The third-level agenda-setting study: an examination of media, implicit, and explicit public agendas in china. *Asian journal of communication* 26(4): 319–332.

Cobb R, Ross JK and Ross MH (1976) Agenda building as a comparative political process. *American Political Science Review* 70(1): 126–138. DOI:10.2307/1960328.

Cohen BC (1963) *Press and Foreign Policy*. Princeton University Press. ISBN 9780691075198. URL http://www.jstor.org/stable/j.ctt183q0fp.

Cong T, Chen B and Ming W (2019) Global science discussed in local social media: Wechat and its comparison with web of science. In: *2019 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*. IEEE, pp. 370–371.

Conway-Silva BA, Filer CR, Kenski K and Tsetsi E (2018) Reassessing twitter's agenda-building power: An analysis of intermedia agenda-setting effects during the 2016 presidential primary season. *Social Science Computer Review* 36(4): 469–483.

Dai Y, Li Y, Cheng CY, Zhao H and Meng T (2021) Government-led or public-led? chinese policy agenda setting during the covid-19 pandemic. *Journal of Comparative Policy Analysis: Research and Practice* 23(2): 157–175.

Dekker R and Scholten P (2017) Framing the immigration policy agenda: A qualitative comparative analysis of media effects on dutch immigration policies. *The International Journal of Press/Politics* 22(2): 202–222.

Deng L and Liu Y (2018) *Deep learning in natural language processing*. Springer.

Devlin J, Chang MW, Lee K and Toutanova K (2018) Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* .

Funk MJ and McCombs M (2017) Strangers on a theoretical train: Inter-media agenda setting, community structure, and local news coverage. *Journalism Studies* 18(7): 845–865.

Gilardi F, Gessler T, Kubli M and Müller S (2022) Social media and political agenda setting. *Political Communication* 39(1): 39–60.

Grimmer J and Stewart BM (2013) Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis* 21(3): 267–297. DOI:10.1093/pan/mps028.

Guo L and McCombs M (2011) Network agenda setting: A third level of media effects. In: *annual conference of the International Communication Association, Boston, MA*.

Guo L and Vargo C (2020) "fake news" and emerging online media ecosystem: An integrated intermedia agenda-setting analysis of the 2016 us presidential election. *Communication Research* 47(2): 178–200.

Guo L and Vargo CJ (2017) Global intermedia agenda setting: A big data analysis of international news flow. *Journal of Communication* 67(4): 499–520.

Guo L and Zhang Y (2020) Information flow within and across online media platforms: an agenda-setting analysis of rumor diffusion on news websites, weibo, and wechat in china. *Journalism Studies* 21(15): 2176–2195.

Harder RA, Sevenans J and Van Aelst P (2017) Intermedia agenda setting in the social media age: How traditional players dominate the news agenda in election times. *The International Journal of Press/Politics* 22(3): 275–293.

Hemphill L and Schöpke-Gonzalez AM (2020) Two computational models for analyzing political attention in social media. In: *Proceedings of the International AAAI Conference on Web and Social Media*, volume 14. pp. 260–271.

Herrero-Jiménez B, Carratalá A and Berganza R (2018) Violent conflicts and the new mediatization: The impact of social media on the european parliamentary agenda regarding the syrian war. *Communication & Society* 31(3): 141–157.

Hofman JM, Watts DJ, Athey S, Garip F, Griffiths TL, Kleinberg J, Margetts H, Mullainathan S, Salganik MJ, Vazire S et al. (2021) Integrating explanation and prediction in computational social science. *Nature* 595(7866): 181–188.

Jansen AS, Eugster B, Maier M and Adam S (2019) Who drives the agenda: Media or parties? a seven-country comparison in the run-up to the 2014 european parliament elections. *The International Journal of Press/Politics* 24(1): 7–26.

Jiang Q, Cheng Y and Cho SK (2021) Media coverage and public perceptions of the thaad event in china, the united states, and south korea: a cross-national network agenda-setting study. *Chinese Journal of Communication* 14(4): 386–408.

Jonkman JG, Boukes M, Vliegenthart R and Verhoeven P (2020) Buffering negative news: Individual-level effects of company visibility, tone, and pre-existing attitudes on corporate reputation. *Mass Communication and Society* 23(2): 272–296.

Judina D and Platonov K (2018) Measuring agenda setting and public concern in russian social media. In: *International conference on internet science*. Springer, pp. 211–225.

Kawakami A, Umarova K and Mustafaraj E (2020) The media coverage of the 2020 us presidential election candidates through the lens of google's top stories. *Proceedings of the International AAAI Conference on Web and Social Media* 14(1): 868–877. URL https://ojs.aaai.org/index.php/ICWSM/article/view/7352.

Kim Y, Kim Y and Zhou S (2017) Theoretical and methodological trends of agenda-setting theory: A thematic analysis of the last four decades of research. *The agenda setting journal* 1(1): 5–22.

Kingdon JW and Stano E (1984) *Agendas, alternatives, and public policies*, volume 45. Little, Brown Boston.

Koltsova O and Nagornyy O (2019) Redefining media agendas: Topic problematization in online reader comments. *Media and Communication* 7(3): 145–156.

Lazer D, Pentland A, Adamic L, Aral S, Barabasi AL, Brewer D, Christakis N, Contractor N, Fowler J, Gutmann M et al. (2009) Social science. computational social science. *Science (New York, NY)* 323(5915): 721–723.

Lee B, Kim J and Scheufele DA (2016) Agenda setting in the internet age: The reciprocity between online searches and issue salience. *International Journal of Public Opinion Research* 28(3): 440–455.

Luo Y, Burley H, Moe A and Sui M (2019) A meta-analysis of news media's public agenda-setting effects, 1972-2015. *Journalism & Mass Communication Quarterly* 96(1): 150–172.

Luo Y and Harrison TM (2019) How citizen journalists impact the agendas of traditional media and the government policymaking process in china. *Global Media and China* 4(1): 72–93.

Males J and Van Aelst P (2021) Did the blue planet set the agenda for plastic pollution? an explorative study on the influence of a documentary on the public, media and political agendas. *Environmental Communication* 15(1): 40–54.

Maniou TA and Bantimaroudis P (2021) Hybrid salience: Examining the role of traditional and digital media in the rise of the greek radical left. *Journalism* 22(4): 1127–1144.

Markiewitz A and Arendt F (2020) Agenda-setting. *The International Encyclopedia of Media Psychology* : 1–19.

McCombs M (1977) Agenda setting function of mass media. *Public relations review* 3(4): 89–95.

McCombs M (2005) A look at agenda-setting: Past, present and future. *Journalism studies* 6(4): 543–557.

McCombs ME and Shaw DL (1972) The agenda-setting function of mass media. *Public opinion quarterly* 36(2): 176–187.

McCombs ME and Shaw DL (1993) The evolution of agenda-setting research: Twenty-five years in the marketplace of ideas. *Journal of communication* 43(2): 58–67.

McCombs ME, Shaw DL and Weaver DH (2014) New directions in agenda-setting theory and research. *Mass communication and society* 17(6): 781–802.

Omoya Y, Akashi J and Kaigo M (2020) Data visualization of texts in the transitions of framing indochinese refugees by japanese television documentaries. *Quality & Quantity* 54(4): 1363–1384.

Pinto S, Albanese F, Dorso CO and Balenzuela P (2019) Quantifying time-dependent media agenda and public opinion by topic modeling. *Physica A: Statistical Mechanics and Its Applications* 524: 614–624.

Platonov K and Svetlov K (2021) Politics-related online communities: Thematic landscape and (para) linguistic features. In: *2021 28th Conference of Open Innovations Association (FRUCT)*. IEEE, pp. 358–364.

Princen S (2007) Agenda-setting in the european union: a theoretical exploration and agenda for research. *Journal of European Public Policy* 14(1): 21–38.

Rogers EM, Dearing JW and Bregman D (1993) The anatomy of agenda-setting research. *Journal of communication* 43(2): 68–84.

Rogstad I (2016) Is twitter just rehashing? intermedia agenda setting between twitter and mainstream media. *Journal of Information Technology & Politics* 13(2): 142–158.

Rosino ML and Hughey MW (2017) Speaking through silence: Racial discourse and identity construction in mass-mediated debates on the "war on drugs". *Social Currents* 4(3): 246–264.

Russell Neuman W, Guggenheim L, Mo Jang Sa and Bae SY (2014) The dynamics of public attention: Agenda-setting theory meets big data. *Journal of communication* 64(2): 193–214.

Scheufele DA and Tewksbury D (2007) Framing, agenda setting, and priming: The evolution of three media effects models. *Journal of communication* 57(1): 9–20.

Shi W, Chen C, Xiong J and Fu H (2019) What framework promotes saliency of climate change issues on online public agenda: A quantitative study of online knowledge community quora. *Sustainability* 11(6): 1619.

Shoemaker PJ and Vos T (2009) *Gatekeeping theory*. Routledge.

Sintes-Olivella M, Franch P, Yeste-Piquer E and Zilles K (2022) Europe abhors donald trump: The opinion on the 2020 u.s. presidential elections and their candidates in the european newspapers. *American Behavioral Scientist* 66(1): 61–85. DOI:10.1177/00027642211005534. URL https://doi.org/10.1177/00027642211005534.

Su Y, Hu J and Lee DKL (2020) Delineating the transnational network agenda-setting model of mainstream newspapers and twitter: A machine-learning approach. *Journalism Studies* 21(15): 2113–2134.

Su Y and Xiao X (2021) Mapping the intermedia agenda setting (ias) literature: Current trajectories and future directions. *The Agenda Setting Journal* 5(1): 56–83.

Sun T and Zhong B (2020) A tale of four cities: A semantic analysis comparing the newspaper coverage of air pollution in hong kong, london, pittsburgh, and tianjin from 2014 to 2017. *Newspaper Research Journal* 41(1): 37–52.

Towner T and Muñoz CL (2020) Instagramming issues: Agenda setting during the 2016 presidential campaign. *Social Media+ Society* 6(3): 2056305120940803.

Tran H (2013) Online agenda setting: A new frontier for theory development. inagenda setting in a 2.0 world: New agendas in communication.

Tran H (2014) Online agenda setting: A new frontier for theory development. *Agenda Setting in a* 2: 205–229.

Van Den Heijkant L, Van Selm M, Hellsten I and Vliegenthart R (2019) Intermedia agenda-setting in a policy reform debate. *International Journal of Communication* 13: 23.

Vonbun R, Königslöw KKv and Schoenbach K (2016) Intermedia agenda-setting in a multimedia news environment. *Journalism* 17(8): 1054–1073.

Wagner P and Payne D (2017) Trends, frames and discourse networks: Analysing the coverage of climate change in irish newspapers. *Irish Journal of Sociology* 25(1): 5–28.

Wang X, Chen L, Shi J and Tang H (2021) Who sets the agenda? the dynamic agenda setting of the wildlife issue on social media. *Environmental Communication* : 1–18.

Wanta W and Ghanem S (2007) Effects of agenda setting. *Mass media effects research: Advances through meta-analysis* : 37–51.

Weaver DH (2007) Thoughts on agenda setting, framing, and priming. *Journal of communication* 57(1): 142–147.

Weimann G and Brosius HB (2017) Redirecting the agenda: Agenda-setting in the online era. *The Agenda Setting Journal* 1(1): 63–102.

Weimann-Saks D, Ariel Y, Malka V and Avidar R (2016) Trends in public and media agenda-setting during the 2015 israeli elections. *Israel Affairs* 22(3-4): 727–742.

Wu HD and Guo L (2020) Beyond salience transmission: Linking agenda networks between media and voters. *Communication Research* 47(7): 1010–1033.

Yüksel E and Dingin AE (2021) The pictures in our heads of certain diseases. *The Agenda Setting Journal* 5(1): 8–30. DOI:https://doi.org/10.1075/asj.19012.yuk. URL https://www.jbe-platform.com/content/journals/10.1075/asj.19012.yuk.