

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

Keywords: agenda setting, computational social science methods, big data, literature review

Extended Abstract

The news is increasingly being produced and read digitally, and machine learning and natural language processing algorithms are opening new possibilities to analyse these 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’s seminal study [2] 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 salient for the public opinion is referred to as first-level agenda setting. McCombs [1] later extended agenda setting theory by developing the concept of an attribute agenda. He 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 [4, 6].

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 et al. [3] 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 one-way 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 [5].

Research questions

We conducted a systematic literature review examining the impact computational methods and large-scale data are having on the field of agenda setting. Our goal is not to replicate or update literature reviews mentioned in the previous section, but rather 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 two 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? How do computational approaches inform the study of media agendas at scale?

Data collection and coding process

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 started the systematic review process by developing a coding scheme to identify publications that either used computational approach or analysed large-scale data in agenda setting research. Two independent coders applied the coding scheme to the empirical papers collected in the literature dataset. There are 163 papers in total meeting the screening criteria and included in this review. For the purpose of word limit, we will not describe qualitative interpretations in the extended abstract.

Results

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. 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 1). We are also interested in the increased or decreased importance of keywords in the abstract of agenda setting and big data research across the six years. 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 (See Figure 2). 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 the “agenda setting” theory. Our study delves into media agenda in the era of big data based on previous research, it indicated that over time, agenda-setting research has expanded in scope, topics, methods, as well as techniques.

References

- [1] Maxwell McCombs. Agenda setting function of mass media. *Public relations review*, 3(4):89–95, 1977.
- [2] Maxwell E McCombs and Donald L Shaw. The agenda-setting function of mass media. *Public opinion quarterly*, 36(2):176–187, 1972.
- [3] W Russell Neuman, Lauren Guggenheim, Set al Mo Jang, and Soo Young Bae. The dynamics of public attention: Agenda-setting theory meets big data. *Journal of communication*, 64(2):193–214, 2014.
- [4] Dietram A Scheufele and David Tewksbury. Framing, agenda setting, and priming: The evolution of three media effects models. *Journal of communication*, 57(1):9–20, 2007.
- [5] Hai Tran. Online agenda setting: A new frontier for theory development. *Agenda Setting in a*, 2:205–229, 2014.
- [6] David H Weaver. Thoughts on agenda setting, framing, and priming. *Journal of communication*, 57(1):142–147, 2007.

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

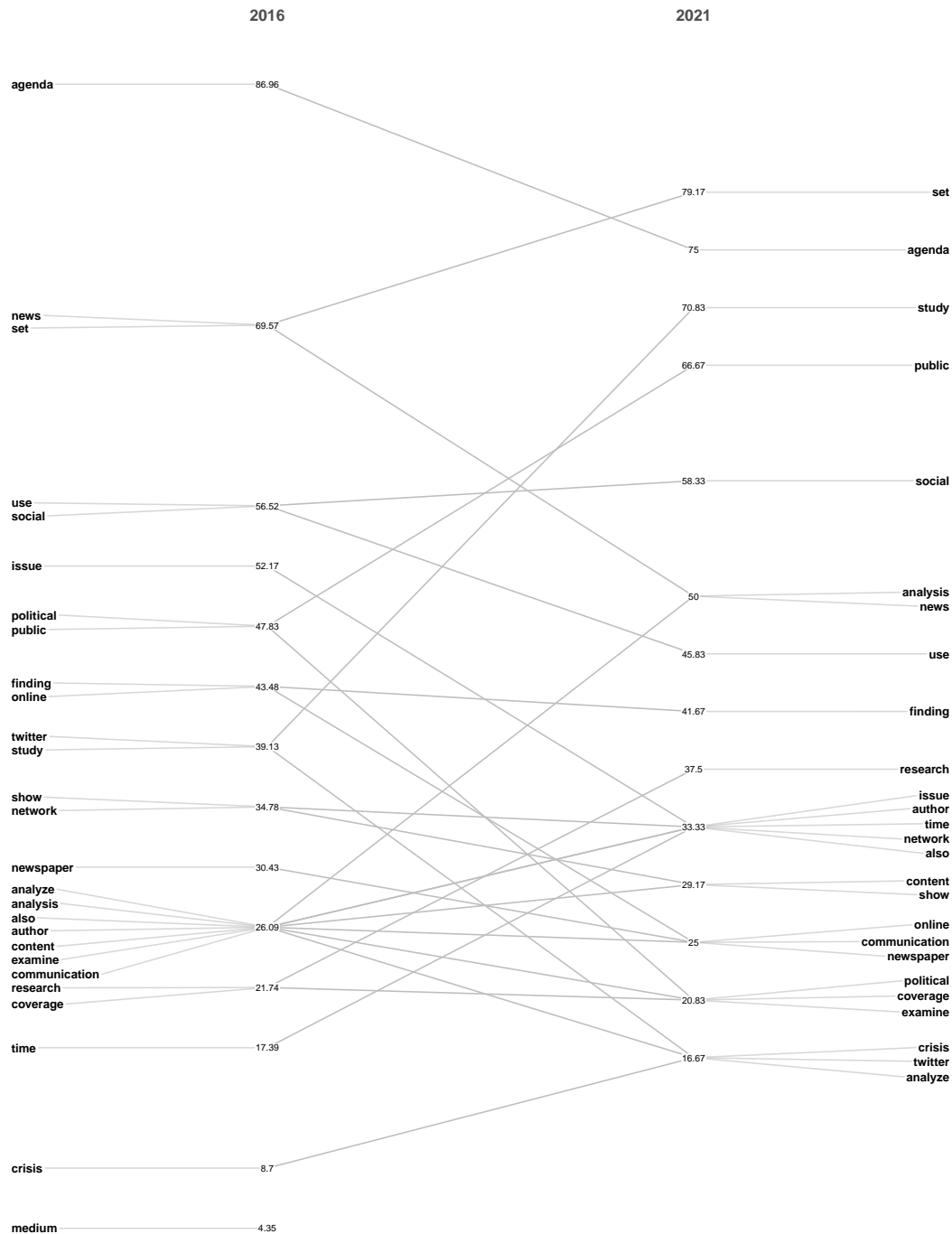


Figure 2: Slope chart of top 20 keywords between 2016 and 2021