Communication Methods and Measures

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/hcms20

What We Can Do and Cannot Do with Topic Modeling: A Systematic Review

Yingying Chen, Zhao Peng, Sei-Hill Kim & Chang Won Choi

COMMUNICATION METHODS AND MEASURES
2023, VOL. 17, NO. 2, 111–130
https://doi.org/10.1080/19312458.2023.2167965

Routledge
Taylor & Francis Group

Check for updates

What We Can Do and Cannot Do with Topic Modeling: A Systematic Review

Yingying Chen [a], Zhao Peng [b], Sei-Hill Kim [c], and Chang Won Choi [d]

[a]School of Journalism and Communication, Renmin University of China, Beijing, China; [b]School of Communication, Emerson College, Massachusetts, USA; [c]School of Journalism and Mass Communications, University of South Carolina, South Carolina, USA; [d]School of Journalism and New Media, University of Mississippi, Mississippi, USA

ABSTRACT

Topic modeling has become an effective tool for communication scholars to explore large amounts of texts. However, empirical studies applying topic modeling often face the critical question of making meaningful theoretical contributions. In this study, we highlighted the importance of theoretical underpinning, the research design, and the methodological details of topic modeling studies. We summarized five normative arguments that address critical issues in theory building and testing, research design, and reliability and validity assessments. Using these normative arguments as criteria, we systematically reviewed 105 communication studies that applied topic modeling. We identified gaps and missed opportunities in previous studies and discussed potential pitfalls for the field.

Increasing communication studies have applied topic modeling to analyze texts from digital media. As an unsupervised machine learning method, topic modeling helps researchers identify latent structures within a large volume of texts (DiMaggio et al., 2013). The method not only provides a lens for scholars to identify new concepts, but also serves as an inductive analysis tool to generate potential hypotheses (Grimmer et al., 2021; Margolin, 2019). As the text data largely unexplored before digital media became available to scholars, topic modeling seems to represent an effective and innovative tool for many communication researchers (van Atteveldt & Peng, 2018).

Despite the methodological strength, empirical studies applying topic modeling face a critical question: what are meaningful theoretical contributions? It is perhaps convenient for communication researchers to employ topic modeling to describe themes in a massive amount of text data without having much prior knowledge. A simple inductive research design seems prevalent in topic modeling studies, despite its potential to propose or test a causal relationship. Topic modeling thus gives readers the impression of providing “a mere descriptive exploration of the corpus when employed on its own” (Törnberg & Törnberg, 2016, p. 418). Moreover, procedures for performing topic modeling have not been fully standardized yet (Maier et al., 2018). The lack of reporting methodological details in previous studies makes it difficult to validate the reported findings and adequately replicate the same findings. As many easy-to-use software packages become accessible, topic modeling can be overused without making theory-informed reasoning or close validation.

To understand how communication researchers have utilized topic modeling, we systematically reviewed 105 empirical studies published in major communication journals from 2009 to 2021. We focus on three challenges that often appear in critiques of topic modeling studies: (1) theory building and testing, (2) research design, and (3) assessments of reliability and validity. We first summarize normative arguments from the current literature in topic modeling and computational social science

CONTACT Yingying Chen yychen1911@hotmail.com Renmin University of China, the School of Journalism and Communication, Mingde Building, 59 Zhongguancun Street, Haidian District, Beijing 100872, China

to address critical issues in each challenge. Then, we use the normative arguments as criteria and conduct a quantitative content analysis of the communication studies using topic modeling. While previous reviews of topic modeling studies focused on post hoc validations of extracted topics (e.g. Maier et al., 2018; Ying et al., 2021), our study attempts not only to review such methodological details but also to highlight the importance of theoretical underpinning and research design of topic modeling studies, which together can enhance their theoretical contributions.

All in all, we intend to show that it takes multiple steps for a study to produce meaningful theoretical contributions using topic modeling. It is not the goal of our systematic review to make a judgment about the value of previous studies. Instead, our goal is to identify critical gaps and missed opportunities in previous research, which will help communication researchers take full advantage of topic modeling and make more useful methodological choices. In particular, it seems highly important to continue the dialogue about the norms and expectations of using topic modeling and other computational text analysis methods properly at this relatively early stage of adopting the methodology (Baden et al., 2022; Grimmer et al., 2022; Maier et al., 2018; Margolin, 2019).

Basic introduction and the strength

Topic modeling is a broad term for computer algorithms that automatically identify latent structures from a large volume of text data. As a popular form of topic modeling,[1] probabilistic topic models such as latent Dirichlet allocation (LDA) estimate the structural patterns in text generation processes, the correlation between themes, and their changes over time based on word occurrence (Blei, 2012). As an abundance of texts from digital media became accessible, communication scholars and other social scientists have used topic modeling as an automatic text analysis tool (see Blei (2012) for more technical introductions of probabilistic topic models).

Topic modeling can be a powerful tool in social science in several aspects. First, it helps researchers gain a quick overview of the major contents from a large volume of text data (DiMaggio et al., 2013; Nelson, 2020). As a dimension-reduction technique, topic modeling transforms a large sample of text into a much smaller set of topics. Second, topic modeling provides a new lens for scholars to identify patterns that would otherwise be undetectable with manual coding alone from a massive amount of texts (DiMaggio et al., 2013). The method can be used as an inductive tool to identify categories that have been largely undiscovered before (Nelson, 2020). Lastly, topic modeling helps communication scholars extract certain meanings from the text data (DiMaggio et al., 2013; Grimmer & Stewart, 2013). Since many theoretical concepts in social science are not directly observable, social scientists have relied on topic modeling as a "text-as-data" method to derive measures of unobservable concepts (e.g., radical rhetoric) from written texts (Ying et al., 2021).

To conduct a topic model, researchers need to preprocess text data and transform it into a matrix (e.g., a document-term matrix) as the model input. Then, researchers pick the best number of topics for a topic model. The primary outputs of topic modeling are (1) topics, each linking to a group of words with higher probabilities than others based on their co-occurrence across documents, and (2) topic proportions of each identified topic in each document. Figure 1 provides an example of the topics and topic probabilities. The words and topic probabilities do not directly reveal the meaning of each topic. Researchers need to interpret and label each topic by carefully inspecting words and documents that are most relevant to each topic. As the procedure of conducting a topic model includes several choices that are seemingly subjective (e.g., the number of topics, the interpretation of topics), computational social scientists have developed a set of norms and expectations for how to use topic modeling properly and why. Being well-informed about the norms and expectations will help communication scholars make better decisions in designing and executing a topic modeling study.

[1]Other non-negative matrix factorization (NMF)-based models (Shi et al., 2018) are also considered a topic modeling technique, which researchers can use to classify documents.

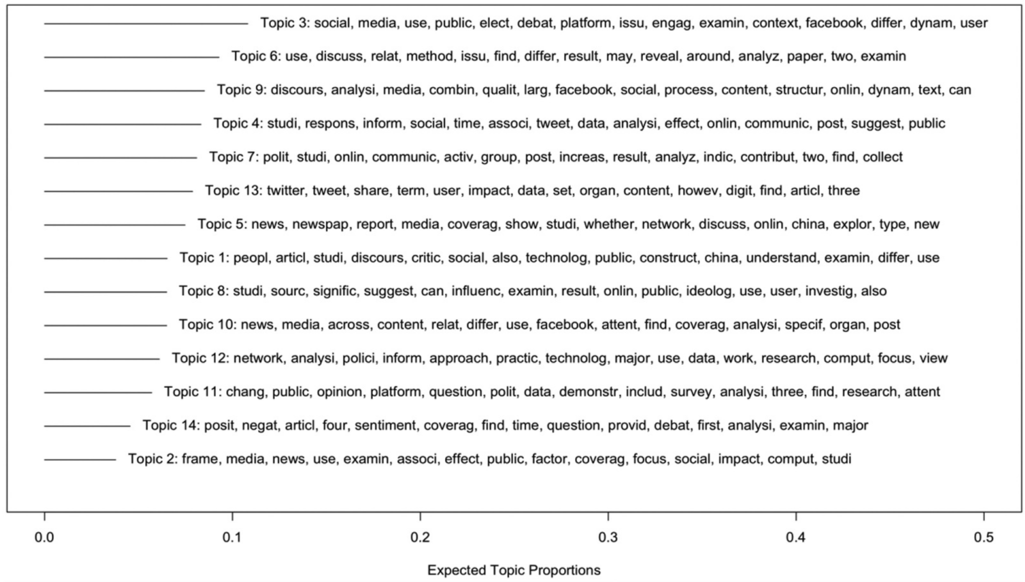

Figure 1. A Visualization of Topic Modeling Output. *Notes*. Figure 1 shows an example of the output from a structural topic model (K=14). The model input is the abstract of the empirical studies using topic modeling in communication.

Theory building and testing

Normative argument 1: proper conceptualization

A fundamental expectation in theory building and testing is that a topic modeling study needs theory-based conceptualizations in the research context. In other words, researchers have theories to explain the reasons for using topic modeling to analyze texts. We should first clarify the meaning of "theory" in computational social science (CSS) studies. A part of theory-driven research involves generating hypotheses based on a formal theory (particularly the meso-level theory, such as agenda setting) and using data to test the hypotheses. In CSS studies, a theory can denote specific patterns of previous findings or a theoretical framework that can explain what affects how digital trace data are created, distributed, and consumed (Patty & Penn, 2014). Similarly, Margolin (2019) considers a "theory" in CSS studies as causal claims (A may cause B, or why/how A may cause B). Waldherr et al. (2021) suggest that CSS researchers connect their research to macro-level theories, such as the theory of mediatization (Krotz, 2007) and the theories of the public sphere (Habermas, 1991). Macro-level theories can provide normative frameworks and concepts that may better explain the complexities, interdependence, and multi-level dynamics that are common for communication phenomena in the digital world (Schroeder, 2018; Waldherr et al., 2021).

Topic modeling studies often show limited connections to theories while mainly describing the frequencies and nature of certain contents. Admittedly, topic modeling was created for information retrieval but not necessarily for testing or building social science theories (Ying et al., 2021). Topic modeling produces data-driven outcomes and assumes that the prior knowledge of a given text is limited. However, this does not mean that the application of topic modeling is limited only to largely descriptive studies. Before employing topic modeling, researchers can contextualize a specific communication phenomenon within a theoretical framework. After running topic modeling, researchers can use specific theoretical lenses to interpret their findings. By incorporating theoretical contexts into a topic modeling study, researchers can examine potential predictors and outcomes of topics in digital media.

A proper conceptualization of communication phenomena of interest could be an important start for theory building and testing (Slater & Gleason, 2012). To set a research context in theories, researchers can conceptualize the problem of interest, explain critical concepts, and generate theory-guided research questions. Digital-trace behavioral data provide opportunities for communication scholars to identify new concepts or refine an existing concept. Communication scholars can also conceptualize a data-generating process in a theoretical framework instead of treating it as a concrete case study. For example, studies examining the formation of public opinion on social media may tie the online exchange of opinions to “networked public spheres” (Waldherr et al., 2021, p. 164). Likewise, Studies analyzing social media engagement metrics and their effects on election results may broadly consider social media engagement metrics as a type of “politically oriented collective behavior” (Margolin, 2019, p. 12). Considering the importance of conceptualization in theory building and testing, we first ask:

RQ1: To what extent have the previous topic modeling studies incorporated a theory-based concept-(s) in the literature review?

Normative argument 2: test hypotheses or generate new hypotheses

Social scientists often tend to move beyond exploration and make inferences about the causes or outcomes of text data (Roberts et al., 2016). As a critical part of theory building and theory testing, another expectation is that researchers can use topic modeling deductively to test hypotheses based on prior literature findings (theories). A deductive study starts with known propositions for hypothesis testing based on a theory that explains a communication phenomenon. To utilize topic modeling to test hypotheses, researchers can propose the presence of specific topics in a given text or compare their prominences across the texts from different sources, regions, or time periods. For instance, researchers can use structural topic modeling (STM) (Roberts et al., 2016) to test how topic probabilities covariate with variables from the text metadata (e.g., publishing time, authors, sources). Researchers can also propose the potential causes or effects of topics. Topic modeling can process a large quantity of data. This strength is helpful in testing hypotheses by analyzing a massive volume and diversity of digital data.

Alternatively, researchers can use topic modeling inductively to generate new hypotheses based on the interpretations of topics. Topic modeling is a powerful tool for inductive studies, but “inductive” differs from “descriptive.” An inductive study begins with an explorative theory-guided analysis and concludes with propositions for generating new hypotheses. Describing the nature of texts is only a part of an inductive study. The inductive approach is typical for qualitative text analysis studies. For example, topic modeling is a critical part of the computational grounded theory approach (Nelson, 2020). Researchers first use topic modeling to identify patterns in an inductive analysis, interpret the patterns based on the prior literature, and then generate or refine new hypotheses. In this context, an inductive or exploratory analysis using topic modeling should ground in theories and contribute to the current literature.

Testing or proposing new hypotheses in a study may indicate that the study is theory-based, either testing an existing theory or building a new one (Slater & Gleason, 2012). To examine the extent to which topic modeling studies have linked their findings to theory testing and building, we put forward the following research questions:

RQ2: To what extent have studies integrated topic modeling results to test hypotheses?

RQ3: What communication theories or concepts have the reviewed studies tested using topic modeling results?

RQ4: To what extent have communication studies only used topic modeling to describe the content of texts? To what extent do those studies generate new hypotheses for future research?

Normative argument 3: pick the most appropriate method

The third expectation is that topic modeling is not always the best method to extract every concept of interest and researchers need to pick the most appropriate text analysis method. Admittedly, as another aspect of theory building, topic modeling provides innovative and scalable ways to extract meaningful theoretical concepts from text data. For example, in political science studies, researchers utilize the method to measure existing concepts, such as frames (DiMaggio et al., 2013) and political agendas (Grimmer & Stewart, 2013). In a systematic review of management studies, Hannigan et al. (2019) found that management research also used topic modeling to explore new constructs and advance conceptualization. However, topics are simply clusters of words generated by algorithms. The estimation of topics relies on the assumptions of a topic modeling algorithm. For example, critical assumptions about LDA and many of its extensions include that (1) words that are semantically similar tend to cluster together in a topic; (2) the orders of words in a document can be neglected to understand the major content in a document (the "bag-of-words" assumption); (3) each document is a representation of a set of topics and each topic is a representation of a set of words from the text data (Blei, 2012; Blei et al., 2003). A topic model produces inaccurate results when used in violation of assumptions. Thus, researchers need to consider other more sophisticated methods (e.g., supervised machine learning, deep learning, embedding techniques) or rely on manual content analysis (Baden et al., 2022; Nelson et al., 2018) for more accurate results.

Topic modeling may not generate the most accurate results in three conditions. First, topic modeling cannot accurately identify implicit concepts in texts (Grimmer et al., 2022). The bag-of-words assumption and the proportions of word co-occurrence constrain the performance of many topic models. Therefore, if the presentation of a concept depends on the syntactic relationship between words, orders of words, or linguistic features that only occasionally cluster together in texts, topic modeling may not be the best choice for identifying the concept. For example, Nelson et al. (2018) employed topic modeling to identify news reports that covered economic inequality. They found that topic modeling could not detect news that mentioned economic inequality as a minor theme and news that covered implicit inequality (e.g., discussion about low-income workers vs. good-fortune executives). In comparison, supervised machine learning methods produced better results. Manual content analysis based on a sufficient sample is perhaps more reliable than topic modeling to identify implicit concepts from texts (Baden et al., 2022). More recent innovations combining transformer models also address the limitations of the bag-of-words assumption. Particularly, topic models combining Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al., 2019) or Sentence BERT (Reimers & Gurevych, 2019), such as BERTopic (Grootendorst, 2022) consider the context of words in a document using document-embedding techniques.

Second, topic models assume that semantic words co-occur together and form a topic, but this assumption does not hold for multilingual texts (Maier et al., 2022). To estimate topics more accurately, researchers need to rely on multilingual probabilistic topic modeling, such as bilingual LDA (Vulić et al., 2015), or rely on machine translation, multilingual dictionary, or conduct different text processing procedures using language-specific stop words and lemmatization (Lind et al., 2021; Maier et al., 2022). Researchers also developed language-specific models considering the unique linguistic patterns of different languages. For example, (Q. Zhao et al., 2011) developed the character-word topic model to analyze Chinese.

Lastly, topic models (e.g., LDA, STM) assume that a document is a representation of multiple topics. This assumption affects the accuracy of analyzing short texts (e.g., tweets), particularly for texts shorter than 50 words (Vayansky & Kumar, 2020). For example, applying an LDA to analyze short texts will pose a data sparsity problem, as the frequency of words in an individual text makes it difficult to determine the precise correlation between words (Vayansky & Kumar, 2020). To address this limitation, researchers aggregate short texts into a larger corpus to avoid the sparsity problem (Albalawi et al., 2020) or use topic models for short texts, such as the biterm topic model (X. Yan et al., 2013).

Our intention here is not to urge researchers to find a model that can identify a ground-truth structure in texts but to help researchers become aware of the limitations of topic modeling. Along with other textual analysis methods, topic modeling estimates one of the many possible structures in texts (Grimmer et al., 2022; Mohr & Bogdanov, 2013). Besides, the previous literature has highlighted the importance of conducting *post hoc* tests to ensure the validity of topic modeling results, that is, whether the results are relevant to a specific research context (Grimmer & Stewart, 2013; Maier et al., 2018; Ying et al., 2021). In addition, we expect that before running a topic model, researchers consider why topic modeling over other methods is the most relevant to a specific research context under investigation. Therefore, we ask:

RQ5: What concepts have been identified using the output of topic modeling in the previous communication studies?

RQ6: What are the topic modeling algorithms used in the previous communication studies?

RQ7: What topic modeling techniques did the previous studies use to analyze non-English, multilingual, or short-text data?

RQ8: How did the previous studies explain using topic modeling over another method?

Research design

Normative argument 4: mixed method design

We expect that researchers can integrate topic modeling with other data analysis methods or other types of data in a mixed-method design. Research design is a plan to use data and methods to find empirical evidence for a social inference (King et al., 1994). We consider two possible ways of using topic modeling inductively and deductively in a mixed-method design. The first is to combine topic modeling with other data analysis methods. For example, qualitative researchers often interpret topics using discourse, word, or narrative analysis (Isoaho et al., 2021). To compare the presence of topics between different sources of texts, quantitative researchers can use statistical inference methods (e.g., regression, Chi-squared tests). The second way is to merge texts with other types of data (quantitative or qualitative). For example, researchers can merge text data with the document meta-data (e.g., authors, sources, publishing dates), social media metrics (e.g., shares, retweets), video, or image data from social media. Researchers can also synthesize aggregated data from Google's community mobility, public opinion poll, voting, or presidential approval data. Synthesizing text data with other forms of data allows researchers to establish a relationship between topics and other variables by treating content data (e.g., topics) as potential treatment (i.e., the cause) or outcome (i.e., the effect).

The relation between topic modeling and other data collection or analysis methods could be parallel or sequential. In parallel mixed designs, multiple analyses may co-occur or in lapsed parallel strands to address different aspects of the same questions; in a sequential mixed method design, researchers apply multiple data analysis methods in consecutive order (Teddlie & Tashakkori, 2010). Topic modeling and other data analysis methods complement each other and make the entire study robust. Regarding mixed method designs, we ask:

RQ9: To what extent have communication studies integrated topic modeling with other data analysis methods and what were such methods?

RQ10: To what extent have communication studies synthesized topic modeling output with other types of data and what was such data?

Reliability and validity

Normative argument 5: report the assessments of reliability and validity

We expect that previous topic modeling studies report the reliability and validity assessments. Since topic modeling always gives a result, its findings can be misleading without rigorous reliability and validity checks (Grimmer & Stewart, 2013). The reliability check evaluates whether the topics are reproducible in new datasets; the validity check evaluates whether the topics represent the actual latent structure of the texts and shows construct validity (Quinn et al., 2010). The construct validity tests are particularly critical for studies that use topics as a proxy measurement of a concept. However, previous reviews of topic modeling studies (Maier et al., 2018; Ying et al., 2021) show that not all social science research utilizing topic modeling has consistently reported reliability and validity checks. Maier et al. (2018) suggested standards for testing as well as reporting reliability and validity in topic modeling studies. Their standards advocated for researchers to (1) specify data preprocessing procedures and procedures for determining the number of topics, (2) conduct reliability checks, and (3) conduct validity checks.

Both data preprocessing procedures and the number of topics influence the stability of topic modeling results. For example, (Denny & Spirling, 2018) found that topic models with the same number of topics will produce different sets of topics if the text input is preprocessed differently (e.g., stemming or without stemming). In addition, a larger topic number tends to produce more specific topics at the risk of creating "overly fine subcategories," while a smaller topic number may generate categories that mix different subcategories in a large "amorphous" topic (Quinn et al., 2010, p. 216). Therefore, transparency in the text preprocessing and topic number selection procedures helps other researchers understand the reliability and validity of the extracted topics (Maier et al., 2018).

Maier et al. (2018) found that researchers often take two approaches in reliability assessments. In one approach, they estimate multiple topic models with varying numbers of topics, manually label the topics, and then inspect whether similar topics appear across different models. In the other approach, researchers split the complete text data into training and testing data. The training data are for building a proper topic model. The testing data are for replicating the model and assessing whether similar topics appear. Researchers can also conduct supervised machine learning in testing data to confirm the prevalence of the topics identified by topic modeling (Nelson, 2020).

In terms of validity checks, previous studies proposed three types of validity checks (Grimmer & Stewart, 2013; Quinn et al., 2010). The first type is face validity, which examines the validity *within* a topic and refers to the validity of topic interpretation. Specifically, when interpreting extracted topics, human coders need to read the most representative words (the highest probability or the most exclusive words) and the most representative documents of a topic. At least two coders label the topics independently and agree upon the labeling to ensure that the labels reflect the true meaning of the extracted topics.

The second type of validity is semantic validity, which indicates the extent to which the underlying words within a topic share a common substantive meaning but show a meaningful difference from the words that belong to other topics. One way to check semantic validity is to manually inspect the words and documents related to a topic and determine whether the topic has a distinctive meaning. Only referring to metrics that measure a model's predictive quality (e.g., held-out likelihood) is insufficient to tell the semantic validity, as the high predictive quality does not lead to interpretable topics (Chang et al., 2009).

The third type of validity check relates to construct validity. When using topic modeling to extract a concept from text data, it is critical to have extra validation to ensure that topics reflect the concept of interest (Ying et al., 2021). One way to evaluate the construct validity is to examine the correlation between the extracted topics and other measurements of the same concept (Quinn et al., 2010). For instance, Jacobi et al. (2016) performed an LDA topic model to identify news frames about nuclear technology and benchmarked the results with manual coding from a previous study (Gamson & Modigliani, 1989). To assess the extent to which topic modeling studies have followed standards proposed by previous studies (Grimmer & Stewart, 2013; Maier et al., 2018; Quinn et al., 2010), we ask the following research questions:

RQ11: To what extent have communication studies applying topic modeling reported the procedures of preprocessing the data and determining the topic numbers?

RQ12: To what extent have communication studies applying topic modeling reported validity and reliability checks and how?

Materials and method

We selected all 89 peer-reviewed communication journals in English from the 2019 Edition of Journal Citation Reports retrieved from the Web of Science database. We first searched communication studies published from January 2009 to May 2021[2] (the most recent month before our search). We identified 201 studies that mentioned "topic model" or "topic modeling" in the title, abstract, keywords, or the main text from the Web of Science database and journal publishers' websites. Our systematic review included only empirical studies[3] ($n = 105$), excluding methodological and conceptual papers. Table 1 shows all the journals and the number of empirical studies included in our review.

We set the five normative arguments as the criteria to analyze 105 empirical studies regarding theory building and testing, research design, and reliability and validity assessments. In theory building and testing, we first labeled the key concepts that a study incorporated in the literature review (RQ1). Then, we quantitatively coded whether a study proposed and tested any hypotheses (RQ2) and the relevant communication theories (RQ3). Next, we coded whether a study used topic modeling only to describe the content and whether the study generated new hypotheses for future research (RQ4). Lastly, we labeled what concepts that topics measured (RQ5), the topic modeling algorithms (RQ6), sources of texts (e.g., Facebook posts, tweets, news reports), text languages (RQ7), and whether a study explained why using topic modeling instead of other methods (RQ8).

In research designs, we first coded whether other data analysis methods (e.g., network analysis, sentiment analysis, regression) were integrated with topic modeling and labeled the data analysis methods (RQ9). We also coded whether other forms of data (e.g., survey data, experimental data, election data) were linked to topic modeling results and coded the types of data (RQ10).

In the reliability and validity assessment, we coded whether each study described the text preprocessing steps and the procedures for determining the number of topics (RQ11). We then coded whether each study reported reliability and validity checks and explained how they were assessed (RQ12).

[2]We selected 2009 as the starting point because in the same year, Lazer et al. (2009) published a pioneer study introducing computational social science.

[3]We excluded studies focusing on methodological innovation (e.g., studies from Communication Methods and Measures) because it is difficult to compare these studies with other empirical studies whose goals are typically theory building and theory testing. However, we still intergrade important topic modeling studies in the literature review and discussion.

Table 1. All Journals Ranked by the Number of Empirical Studies.

Journal Title	Empirical studies
Information, Communication & Society	13
Social Media + Society	
Digital Journalism	7
Political Communication	6
Journalism Studies	
Health Communication	
New Media & Society	4
Journal of Information Technology & Politics	
Journal of Computer-Mediated Communication Policy & Internet	3
Public Relations Review	
Telecommunications Policy	
Journal of Broadcasting & Electronic Media	
International Journal of Communication	
Communication Research	2
Journalism	
Journal of Health Communication	
Discourse & Communication	
Discourse & Society	
Games and Culture	
Asian Journal of Communication	
Journal of Advertising, Journal of Communication, International Journal of Advertising, The International Journal of Press/Politics, Mobile Media & Communication, European Journal of Communication, Public Opinion Quarterly, Public Understanding of Science, Journalism & Mass Communication Quarterly, International Communication Gazette, Feminist Media Studies, International Journal of Business Communication, Journal of Language and Social Psychology, Television & New Media	1

Since there were no consistent validity and reliability testing approaches, we summarized the main approaches from the previous studies (Grimmer & Stewart, 2013; Maier et al., 2018; Quinn et al., 2010) for the manual coding (see p. 2–p.3 in the supplementary document). We first tested the initial coding instrument with a subset of studies and revised coding instruments during the coder training and pretesting procedures. In coder training sessions, two coders established an accurate and mutual understanding of each variable in the coding instrument, and a third person resolved coder disagreements. Then, by double-coding a random sample of 22 studies (about 15% of the 105 articles), two coders reached intercoder reliability (Cohen's kappa) ranging between .83 and 1, with an average of .92. The two coders coded the rest of the studies separately.

Results

Theory building and testing

Most studies ($n = 98$, 93%) incorporated a key concept and explained it in the literature review (RQ1). As shown in Table 2, the concepts are primarily from communication theories like agenda setting, framing, or the public sphere theories. For instance, researchers conceptualized social media as a public sphere (X. Zhao et al., 2018) and social media users as networked publics (Yang et al., 2021) or networked news public (Kwon et al., 2019). Other concepts are related to the radical content online, such as populism (e.g., Stier et al., 2017) and far-right movement (Kaiser et al., 2020), and the unintended consequences of using digital media, such as filter bubbles (Bechmann & Nielbo, 2018), cyberbalkanization (Yang et al., 2021), and effects of news recommendation algorithms (Möller et al., 2018). The findings suggest that most studies we reviewed had theory-based conceptualities in the research context.

We examined the extent to which topic modeling studies tested hypotheses or generated new hypotheses (RQ2-RQ4). We found that only 16 studies (15%) tested a hypothesis. Framing and agenda-setting theory ($n = 4$) were the most common communication theories tested in these hypothesis-testing studies (Ghosh et al., 2022; Guo, 2019; Walter & Ophir, 2020; Xu et al., 2020). We also

Table 2. Major Concepts in the Literature Review of the Reviewed Studies.

Category	Examples of Concepts
Agenda setting and framing	news frames (Walter & Ophir, 2020), affective news (Savolainen et al., 2020), network agenda setting (Su et al., 2020), agenda setting (Papadouka et al., 2016), networked framing (Pöyhtäri et al., 2021)
Social media as a public sphere	public frame building (van der Meer, 2018), public discourse (e.g., Pantti et al., 2019), networked publics and cyberbalkanization (Yang et al., 2021), networked news public (Kwon et al., 2019), counter public (Zeng et al., 2020), networked public sphere (X. Zhao et al., 2018)
Radical content online and the unintended consequences of using digital media	• populism (e.g., Stier et al., 2017), anti-establishment speeches (Ceron et al., 2020), far-right movement (Kaiser et al., 2020) • anti-vaccination movement (Smith & Graham, 2019), science skeptics (P. Yan et al., 2021) • Islamophobia and anti-feminism (Al-Rawi et al., 2021), anti-immigration (Merrill & Åkerlund, 2018), racism (Pantti et al., 2019) • online trolling/harassment (Cook et al., 2020; Kargar & Rauchfleisch, 2019) • filter bubbles/echo chamber (and the effects) (Bechmann & Nielbo, 2018); online polarization/homophily Robles et al. (2020); news recommendation algorithms (Möller et al., 2018) • social bots (Assenmacher et al., 2020) • interactive political polarization (Yarchi et al., 2020), online censorship (Zeng et al., 2017)
Digitized communication in selected domains	• News: sources/themes in news coverage (Von Nordheim et al., 2018), news cycles (Segal & Soffer, 2020) • Politicians: online political self-expression (e.g., Koltsova & Shcherbak, 2015) • Digital advocacy: online mobilization and demobilization (e.g., Kligler-Vilenchik et al., 2020), hashtag activism (Lindgren, 2018), hashtag politics (Johnson et al., 2019) • Health: social support (e.g., Erčulj & Žiberna, 2021)

found that among 20 studies (19%) performed topic modeling only to describe text content, six generated hypotheses for future studies in the discussion section (e.g., Assenmacher et al., 2020; Koltsova & Koltcov, 2013). Missed opportunities seem to exist in previous studies – the opportunities to test or generate new hypotheses.

Most studies ($n = 80$, 76%) conceptualized topics simply as topics, and only 25 studies (24%) used topics as a proxy measure of a concept (RQ5). Most typically, topics were conceptualized as themes ($n = 5$), discourses ($n = 6$), and frames ($n = 6$). Other conceptualizations include frame packages (Ghosh et al., 2022), frame elements (Ophir et al., 2021; Walter & Ophir, 2020), narratives (X. Zhao et al., 2018), knowledge domains (Righi et al., 2020), genre (Faisal & Peltoniemi, 2015), issues (Tresch & Feddersen, 2019), and agenda (Koltsova & Koltcov, 2013; Su et al., 2020).

Additionally, most studies (82%) used either LDA ($n = 63$, 60%) or STM ($n = 23$, 22%) (RQ6). Other less used algorithms included non-negative matrix factorization (NMF), agglomerative hierarchical clustering algorithm, latent semantic analysis, correlated topic model (Blei & Lafferty, 2007), document influence model (Gerrish & Blei, 2010), and hierarchical topic model (Grimmer, 2010). Notably, five studies did not mention which topic modeling algorithm they used.

Although English ($n = 69$, 66%) was the primary language of the texts, we still found 35 studies (33%) analyzing non-English texts, such as Chinese ($n = 10$, 10%), German ($n = 6$, 6%), and Hebrew ($n = 5$, 5%) and five studies (5%) analyzed multilingual texts (Table 3). Among the 40 studies (38%) that analyzed non-English texts or multilingual texts LDA ($n = 23$) and STM ($n = 11$) were also the most popular algorithms; only 26 reported language-specific techniques in text preprocessing (RQ7). For example, researchers used Jieba for Chinese word segmentation (Guo, 2019) or MyStem for lemmatizing Russian (Koltsova & Koltcov, 2013). Three multilingual studies reported that they conducted machine translation (Kligler-Vilenchik et al., 2020) or created a separate topic model for each language (Tresch & Feddersen, 2019; Xu et al., 2020). As we examined the sources of text data, we found that tweets ($n = 33$, 31%), news reports ($n = 29$, 28%), and Facebook posts ($n = 12$, 11%) were the most

Table 3. Languages of Texts Data in the Reviewed Studies.

Language	Studies
English	69
Chinese	10
German	6
Hebrew, Swedish	5
Dutch, Finnish, Spanish	3
Russian	2
Arabic, Danish, Italian, Norwegian, Persian, Slovenian	1
Multilingual	5

Table 4. Types of Text Data in the Reviewed Studies.

Texts Type	Studies
Tweets	33
News	29
Facebook posts	12
Responses to open questions in a survey	6
Online forum threads	5
Blogs	4
WeChat articles, Reddit posts	3
Instagram comments, YouTube comments, Weibo, video transcripts, WhatsApp group chats	2
Online news comments, Overboard comments, press releases, Novels, chat logs in online games, program descriptions (from Github, Bitbucket, and GitLab), corporate annual reports, public speech and statement of FTC commissioners, description of video games, website articles, the textual replies to the public consultation, corporate R&D documents	1

popular types of text data (Table 4). Other types of text data (n = 36, 36%) include blogs, novels, Reddit posts, and Instagram comments. We took tweets as an example of short texts and found that 21 of the 33 studies analyzing tweets used LDA, even though aggregating texts or applying topic models for short texts could generate more accurate results.

As we examined how researchers explained why topic modeling was the most appropriate method for the study (RQ8), we found that about half of the studies (n = 49, 47%) only introduced what topic modeling was or illustrated how the model worked but did not explain why it was chosen for the study. Among the studies that provided an explanation, topic modeling became an appropriate method often because it was a convenient tool to analyze a large volume of data automatically. However, it is perhaps misleading to emphasize the convenience of topic modeling without mentioning the importance of theories and human involvement when interpreting topics. In addition, we also observed a few cases of using topic modeling when other methods might generate more accurate results. For example, several studies manually assigned each document to a known category based on the highest topic probability or an arbitrary threshold probability, although supervised machine learning could be a better option. Three studies used topic modeling to identify themes from a relatively smaller sample (<1,000) of news articles, which may not justify the use of topic modeling instead of manual coding. In sum, our findings show that topic models were used in violation of their assumptions in some studies and there was another missed opportunity, that is, applying other methods that could have generated more accurate results than topic modeling.

Research design

Most studies (n = 94, 90%) integrated topic modeling with other data analysis methods in a mixed-method design (RQ9). First, to interpret topics, researchers conducted a qualitative analysis using critical discourse analysis (Lindgren, 2018; Pantti et al., 2019; Wahlström et al., 2020), word analysis (Fonseca et al., 2019), or quantitative analysis using manual coding (Kligler-Vilenchik et al., 2020; Koltsova & Shcherbak, 2015). Second, researchers conducted a follow-up statistical inference to (1) predict changes in topics over time using Autoregressive Integrated Moving Average (ARIMA) models (Ghosh et al., 2022;

Zeng et al., 2017) or Vector Autoregressive (VAR) models (Ophir et al., 2021), and (2) compare topics in texts of different sources using a Chi-squared test (Papadouka et al., 2016), or (3) establish a predictive or potentially explanatory relationship using, for example, a regression (Sweitzer & Shulman, 2018; Van Der Velden et al., 2018; Walter & Ophir, 2020). Lastly, topic modeling was used in parallel with other exploratory analysis methods, such as sentiment analysis (Ceron et al., 2020; Robles et al., 2020; Zeng et al., 2020), hashtag network analysis (Zhang, 2018), or retweet network analysis (Zheng & Shahin, 2020). Topic modeling was also used with principal component analysis (Kaiser et al., 2020) or a topic network analysis (Ophir et al., 2021; Walter & Ophir, 2020) to examine the relationship between topics.

We found 21 studies (20%) that linked topic modeling findings to other data forms (RQ10). The main goal was to examine the potential causes or effects of topics. For example, researchers merged topic probability with voting data (Bright et al., 2020; Koltsova & Shcherbak, 2015; Walter & Ophir, 2020), vaccination rates (Smith & Graham, 2019), Google mobility data (Ophir et al., 2021), air quality data (Gurajala et al., 2019), and public opinion (Tresch & Feddersen, 2019; Van Der Velden et al., 2018). Alternatively, researchers also collected qualitative data from semi-structured interviews (Thorson et al., 2020), ethnography (Lu & Pan, 2020), quasi-experiments (van der Meer, 2018), and surveys (P. Yan et al., 2021) to explain the causes of topics. In sum, these findings suggest that communication researchers have often used a mixed method design to examine potential predictors and consequences of topics instead of simply describing the presence or proportions of specific topics within a given text.

Reliability and validity

We found that only 67 studies (64%) reported text preprocessing steps and 65 studies (62%) (Q11) reported how the number of topics was determined, raising concern about the replicability of the findings in previous topic modeling studies. Furthermore, only 12 studies (11%) reported *both* validity and reliability checks, and as many as 41 studies (39%) reported neither reliability nor validity checks (RQ12). Particularly, studies that applied algorithms other than LDA and STM (e.g., k-means clustering, NMF) mainly did not report reliability and validity checks. The low proportion of studies reporting reliability and validity poses the problem of overfitting.

As many as 52 studies (61%) only conducted validity checks without assessing reliability. Of the 25 studies (24%) that conceptualized topics as a theoretical concept, 18 conducted construct validity by manually inspecting whether topics accurately reflected the concept of interest. Overall, the steps in validity checks are far from being standardized. This finding is consistent with findings from the previous review studies (Maier et al., 2018; Ying et al., 2021), which reported an overall lack of validity check in topic modeling studies. More recent studies in our data reported that they followed the standards in Maier et al. (2018) to perform topic modeling.

Discussion and implications

Communication researchers have shown a growing interest in topic modeling, and its increasing presence in communication studies also raises the question of how to utilize this method to produce more meaningful theoretical contributions. To learn from previous studies, we summarized five normative arguments in (1) theory building and testing, (2) advancing research design, and (3) assessing reliability and validity. Then, using the normative arguments as criteria, we systematically reviewed 105 empirical studies that used topic modeling as part of communication research. The goal of this review is not to judge or challenge the value of specific studies, but to identify the missed opportunities and discuss potential pitfalls for the field.

The strength of topic modeling as an exploratory analysis method

Computational social scientists have acknowledged the strength of topic modeling as a method to perform inductive analysis or extract concepts of theoretical interest. Our findings showed that many communication studies utilized topic modeling more as an inductive method than a method for identifying a concept. Many studies described the strength of topic modeling as an automated method that can analyze a large volume of texts, delineate the major themes, and trace their longitudinal changes. Meanwhile, descriptive analysis was not necessarily the end goal for the reviewed studies. Most studies used topic modeling in a sequential or parallel mixed-method design. They combined topic modeling with other data analysis methods (90%) or merged findings from topic modeling with data from other sources (20%). For example, in the sequential mixed method design, researchers combined topic modeling with statistical inference to compare the presence and frequencies of topics in texts of different contexts (e.g., platforms, culture, time, sources) or identify predictive or explanatory relationships between topics and other variables.

The reviewed studies did not meet some of our normative expectations. While the purpose of using topic modeling can be beyond offering a descriptive analysis, only a few studies proposed a hypothesis ($n = 16$) in the literature review or generated new hypotheses ($n = 6$) based on the interpretation of topics. Most studies applied LDA or STM (82%), even when other methods could generate more accurate results. By violating the assumptions of a topic model, researchers often compromised the accuracy of results. These findings indicated that missed opportunities exist in the field. We also found that the lack of transparency and limited replicability of topic modeling became another vital problem in previous studies. Only a small number of studies (11%) reported reliability and validity assessment, which may seriously challenge the value of topic modeling in communication studies.

The missed opportunities

We identified three missed opportunities in theory building and testing. The first was to use other textual analysis models than LDA or STM, especially when certain characteristics of the texts (e.g., language, length) did not meet the key assumptions of such models. Only a few reviewed studies used alternative options, such as non-negative matrix factorization (NMF), a deterministic topic modeling method to analyze tweets (Bright et al., 2020; Ceron et al., 2020). NMF performed better than LDA in analyzing short texts such as tweets (Chen et al., 2019). Yang et al. (2021) examined the factors that could predict how topics changed over time using the document influence model (Gerrish & Blei, 2010), an extension of dynamic topic modeling considering the sequences of documents (Blei & Lafferty, 2006). Dehghani et al. (2014) used LDA with topic-in-set knowledge – a semi-supervised topic modeling that includes predetermined seed words to constrain the topic estimation process (Andrzejewski & Zhu, 2009). When analyzing multilingual texts, researchers can rely on machine translation and multilingual dictionaries (see Lind et al., 2021; Maier et al., 2022; Reber, 2019). We encourage researchers to provide an explicit justification of why a topic modeling algorithm, instead of other algorithms, is most appropriate for the study and explain how the strength of topic modeling serves the specific goal of the study. Doing so will help researchers evaluate the potential limitations of different text analysis techniques and make the most reasonable methodological choice. For example, in a recent study, Simon et al. (2022) applied BERTopic (Grootendorst, 2022) as a dynamic topic modeling technique to explore how topics change over time.

The second missed opportunity was to utilize topic modeling in testing hypotheses. We recommend that future studies, particularly the topic modeling studies that are quantitatively oriented, explicitly draw hypotheses from the prior knowledge in the field. For example, Margolin (2019) highlighted the importance of testing causal relationships in computational communication studies. A few reviewed studies echo this point with two ways to hypothesize potential causal relationships. First, researchers can start with a widely held explanation of a communication phenomenon and challenge it by proposing an alternative explanation. For example, Yang et al. (2021) hypothesized that inter-community influences would be contingent on the size and stability of online communities, the valence of interactions, and the volume of interactions between online communities. This argument challenged a common explanation of inter-community communication. That is, inter-community communication on social media is polarized and fragmented. They found that online communities were not isolated. Particularly, a large and stable community was more likely to influence the topics of another community; the responses from the target community were also likely to shape the topics of the source community. Second, researchers can compare and assess the validity of several alternative explanations. For instance, Bright et al. (2020) compared broadcast effects (messages from a political party), interactivity effects (the direct interaction between party candidates and the public), and name recognition effects (public mention of names of candidates) as three mechanisms through which political campaigns on Twitter could affect voters. Although the prior literature suggests that political campaign effects come mostly from the direct engagement between politicians and voters, Bright et al. (2020) found that only the broadcast mechanisms produced campaign effects on Twitter.

The third missed opportunity was to generate new hypotheses based on topic interpretation. We recommend that researchers generate a set of explicit propositions, if they intend to use topic modeling primarily in an inductive study. In this case, illustrating the prior knowledge from the literature and having a theoretical lens to interpret the topics lay the foundation for generating new hypotheses. A few studies in our review provided good examples (e.g., Assenmacher et al., 2020; Koltsova & Koltcov, 2013; McKernan, 2019; Rhidenour et al., 2021). Assenmacher et al. (2020) applied topic modeling to identify the capabilities (e.g., chat, gaming capabilities) of social bot programs. The researchers first pointed out that while previous studies paid much attention to the malicious influence of social bots, how widely such sophisticated bots existed in online media remained largely unexamined. Then, they used topic modeling to analyze the description of 45,018 bot repositories – a representative size of bot programs that mainly targeted users on Facebook, Instagram, YouTube, or Twitter. Their findings showed that the capabilities of social bots were too limited to make a malicious influence. They proposed three possible explanations that future studies can validate.

The problems

The lack of reliability and validity assessments in previous studies seems to be a problem rather than a missed opportunity. We identified three general patterns of related problems from our review. The first pattern was conducting no reliability test at all. Instead, we recommend that researchers perform and report a cross-validation check to ensure coherent topic results appear across separate datasets. For instance, Britt and Britt (2021) used five-fold cross-validation and checked whether similar topics would emerge across five randomly sampled text datasets. That is, assessing reliability is a reiterative process to ensure that the patterns of findings identified by topic modeling consistently appear in different data sets. As Grimmer et al. (2022) suggested, when using topic modeling for measurement, researchers first need to split data into a training and a testing set. Then, they analyze the training data to identify the presence of the target concept or the expected patterns of topics. Finally, to ensure the initial findings are reliable and generalizable, researchers need to replicate the initial analysis using testing data. This identification and confirmation process is non-linear and reiterative (Grimmer et al., 2022). Future studies may consider reporting the reiterative reliability assessment in the initial analysis.

The second pattern was to treat topic interpretations or the predictive power of a topic model as a proxy assessment of validity. In many reviewed studies, the processes used for topic interpretation, the methods used to determine the number of topics, or the comparison of parameters that measured predictive quality were often reported as part of validity checks. This created problems for conducting rigorous validity tests. For example, a good model with high predictive power does not necessarily mean its outcomes (i.e., the identified topics) reflect the concepts of interest (Chang et al., 2009). In addition, rigorous validity tests often require additional assessment with human input (Ying et al., 2021). For example, two reviewed studies compared topic modeling results with manual coding (Walter & Ophir, 2020; Yarchi et al., 2020). Therefore, we recommend that future studies refer to the methodological literature (Grimmer & Stewart, 2013; Jacobi et al., 2016; Maier et al., 2018; Ying et al., 2021) that guides validity tests. R packages, such as oolong (Chan and Sältzer (2020) and tosca (Koppers et al., 2021) are tools to test semantic validity.

The last pattern was to use topic modeling to identify documents that belong to a category of interest. In many reviewed studies, researchers assigned a document to a known category (e.g., anti-vaccine opinion, populism, hate/anti-hate speech) based on the highest topic probability or a threshold of topic probability. As Nelson et al. (2018) suggested, the strength of topic modeling is in inductive analysis, and using the method to identify known categories in texts may produce inaccurate results. It is more accurate to use supervised machine learning to identify the documents of interest and then use topic modeling to analyze the themes in the selected documents inductively. However, topic modeling is often a convenient alternative, especially when highly reliable training data for supervised learning is hardly obtainable (Grimmer et al., 2022). In this case, we recommended that researchers not only conduct validity checks of the topics identified in the selected documents but also carefully examine the unselected documents to assess the extent to which documents are falsely unselected (false negative).

Limitations

This review has several limitations. First, we did not systematically examine whether topic modeling was overused in all communication studies that applied computational text analysis methods. However, we still provide evidence of overusing topic modeling by violating the assumptions of a topic model. Second, we did not systematically examine innovative approaches to make causal inferences using text-as-data methods, which is an important future direction for computational social science studies (Grimmer et al., 2022). For example, STM can be a valuable model for making causal inferences by treating text as the outcome in an experiment design (Roberts et al., 2016). However, we still identified research designs to establish potential causal relationships by explaining the generation of topics using semi-conduct interviews (Thorson et al., 2020) or by merging data from other sources as the potential effects of topics (e.g., Bright et al., 2020). Future studies can explore more innovative research designs for making causal inferences utilizing text data.

Conclusion

This systematic review investigates what can or cannot be done with topic modeling to make meaningful theoretical contributions. Researchers need multiple steps in theory building and testing, research design, and reliability and validity assessments. We conclude this review with a list of recommendations. First, researchers need to consider what theories provide a lens (1) to conceptualize the text generation process of interest and (2) to propose theory-driven explanations to the research question. Second, researchers collect training data and conduct an initial analysis using topic modeling. The primary goals are to explore (1) whether and how much the concept or patterns of interest appear in texts and (2) to gain data-driven insights to refine research questions or propose hypotheses. Researchers may also consider integrating topic modeling with statistical inference methods or data from other sources to establish an explanatory relationship between topics and other variables. Third, researchers need to pick the most appropriate text analysis method by comparing the strength of other text analysis methods (e.g., supervised machine learning, manual content analysis) and different topic models. Lastly, researchers need to conduct and report assessments of reliability and validity. Particularly, it is helpful for researchers to (1) split the data into training and testing sets, (2) replicate topic modeling analysis employing testing data as a part of reliability checking, and (3) perform topic modeling and other text analysis methods to analyze training data as a part of validity checking.

Acknowledgement

The authors are grateful to Dr. Tai-Quan 'Winson' Peng, Dr. Marko Bachl, and the three anonymous reviewers for their constructive suggestions.

Disclosure statement

We have no known conflict of interest to disclose.

Notes on contributors

Dr. Yingying Chen (PhD in Information and Media, Michigan State University) is an Assistant Professor in the School of Journalism and Communication at the Renmin University of China. Her research focuses on the intersection of political communication, science communication, information diffusion, and computational social science methods.

Dr. Zhao Peng (Ph.D., Michigan State University) is an Assistant Professor in the Department of Journalism at Emerson College. Her research investigates how communication scientists can more effectively present their findings, including how communication practitioners may apply DEIA knowledge in data visualizations.

Dr. Sei-Hill Kim (PhD, Cornell University) is the Eleanor M. & R. Frank Mundy Professor in the School of Journalism and Mass Communications at the University of South Carolina. His research investigates the role of mass media in shaping public discourse on important issues in politics, public health, and science and technology.

Dr. Chang-Won Choi (PhD, The University of South Carolina) is an Assistant Professor of Integrated Marketing Communications in the School of Journalism and New Media at the University of Mississippi. His research interests include viral advertising, digital advertising, corporate social responsibility (CSR) campaigns, and computational methods for advertising research.

ORCID

Yingying Chen http://orcid.org/0000-0003-4278-8206
Zhao Peng http://orcid.org/0000-0002-2420-5567
Chang Won Choi http://orcid.org/0000-0002-2878-8720

References

Albalawi, R., Yeap, T. H., & Benyoucef, M. (2020). Using topic modeling methods for short-text data: A comparative analysis. *Frontiers in Artificial Intelligence*, *3*(July), 1–14. https://doi.org/10.3389/frai.2020.00042

Al-Rawi, A., Chun, W. H. K., & Amer, S. (2021). Vocal, visible and vulnerable: Female politicians at the intersection of islamophobia, sexism and liberal multiculturalism. *Feminist Media Studies*, 1–18. https://doi.org/10.1080/14680777.2021.1922487

Andrzejewski, D., & Zhu, X. (2009). Latent Dirichlet allocation with topic-in-set knowledge. In Proceedings of the NAACL 2009 Workshop on Semi-supervised Learning for NLP (pp. 43– 48). Stroudsburg, PA: Association for Computational Linguistics.

Assenmacher, D., Clever, L., Frischlich, L., Quandt, T., Trautmann, H., & Grimme, C. (2020). Demystifying social bots: On the intelligence of automated social media actors. *Social Media + Society*, *6*(3), 205630512093926. https://doi.org/10.1177/2056305120939264

Baden, C., Pipal, C., Schoonvelde, M., & van der Velden, M. A. C. G. (2022). Three gaps in computational text analysis methods for social sciences: A research agenda. *Communication Methods and Measures*, *16*(1), 1–18. https://doi.org/10.1080/19312458.2021.2015574

Bechmann, A., & Nielbo, K. L. (2018). Are we exposed to the same "news" in the news feed? *Digital Journalism*, *6*(8), 990–1002. https://doi.org/10.1080/21670811.2018.1510741

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, *55*(4), 77–84. https://doi.org/10.1145/2133806.2133826

Blei, D. M., & Lafferty, J. D. (2006). Dynamic topic models. In *Proceedings of the 23rd international conference on machine learning* (pp. 113–120). New York: Association for Computing Machinery.

Blei, D. M., & Lafferty, J. D. (2007). A correlated topic model of science. *The Annals of Applied Statistics*, *1*(1), 17–35. https://doi.org/10.1214/07-AOAS114

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, *3*(4–5), 993–1022. http://www.crossref.org/deleted_DOI.html

Bright, J., Hale, S., Ganesh, B., Bulovsky, A., Margetts, H., & Howard, P. (2020). Does campaigning on social media make a difference? Evidence from candidate use of twitter during the 2015 and 2017 U.K. elections. *Communication Research*, *47*(7), 988–1009. https://doi.org/10.1177/0093650219872394

Britt, B. C., & Britt, R. K. (2021). From waifus to whales: The evolution of discourse in a mobile game-based competitive community of practice. *Mobile Media and Communication*, *9*(1), 3–29. https://doi.org/10.1177/2050157920934509

Ceron, A., Gandini, A., & Lodetti, P. (2020). Still 'fire in the (full) belly'? Anti-establishment rhetoric before and after government participation. *Information, Communication & Society*, *0*(0), 1–17. https://doi.org/10.1080/1369118X.2020.1776373

Chang, J., Boyd-Graber, J., Gerrish, S., Wang, C., & Blei, D. M. (2009). Reading tea leaves: How humans interpret topic models. *Advances in Neural Information Processing Systems 22 - Proceedings of the 2009 Conference*, Vancouver, British Columbia, Canada, 288–296.

Chan, C., & Sältzer, M. (2020). Oolong: An R package for validating automated content analysis tools. *Journal of Open Source Software*, *5*(55), 12461. https://doi.org/10.21105/joss.02461

Chen, Y., Zhang, H., Liu, R., Ye, Z., & Lin, J. (2019). Experimental explorations on short text topic mining between LDA and NMF based schemes. *Knowledge-Based Systems*, *163*, 1–13. https://doi.org/10.1016/j.knosys.2018.08.011

Cook, C., Conijn, R., Schaafsma, J., & Antheunis, M. (2020). For whom the gamer trolls: A study of trolling interactions in the online gaming context. *Journal of Computer-Mediated Communication*, *24*(6), 293–318. https://doi.org/10.1093/jcmc/zmz014

Dehghani, M., Sagae, K., Sachdeva, S., & Gratch, J. (2014). Analyzing political rhetoric in conservative and liberal weblogs related to the construction of the "ground zero mosque. *Journal of Information Technology & Politics*, *11*(1), 1–14. https://doi.org/10.1080/19331681.2013.826613

Denny, M. J., & Spirling, A. (2018). Text preprocessing for unsupervised learning: Why it matters, when it misleads, and what to do about it. *Political Analysis*, *26*(2), 168–189. https://doi.org/10.1017/pan.2017.44

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, Minneapolis, United States, *1*, 4171–4186.

DiMaggio, P., Nag, M., & Blei, D. (2013). Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of U.S. government arts funding. *Poetics*, *41*(6), 570–606. https://doi.org/10.1016/j.poetic.2013.08.004

Erčulj, V. I., & Žiberna, A. (2021). The role of online social support in patients undergoing infertility treatment–A comparison of pregnant and non-pregnant members. *Health Communication*, *37*(14), 1–7. https://doi.org/10.1080/10410236.2021.1915517

Faisal, A., & Peltoniemi, M. (2015). Establishing video game genres using data-driven modeling and product databases. *Games and Culture*, *13*(1), 20–43. https://doi.org/10.1177/1555412015601541

Fonseca, A. F., Bandyopadhyay, S., Louçã, J., & Manjaly, J. (2019). Caste in the news: A computational analysis of Indian newspapers. *Social Media and Society*, *5*(4), 205630511989605. https://doi.org/10.1177/2056305119896057

Gamson, W., & Modigliani, A. (1989). Media discourse and public opinion on nuclear power: A constructionist. *The American Journal of Sociology*, *95*(1), 1–37. https://doi.org/10.1086/229213

Gerrish, S., & Blei, D. M. (2010). A language-based approach to measuring scholarly impact. *ICML 2010 - Proceedings, 27th International Conference on Machine Learning*, Haifa, Israel, *10*, 375–382.

Ghosh, S., Su, M. -H., Abhishek, A., Suk, J., Tong, C., Kamath, K., Hills, O., Correa, T., Garlough, C., Borah, P., & Shah, D. (2022). Covering #metoo across the news spectrum: Political accusation and public events as drivers of press attention. *The International Journal of Press/Politics*, *27*(1), 158–185. https://doi.org/10.1177/1940161220968081

Grimmer, J. (2010). A Bayesian hierarchical topic model for political texts: Measuring expressed agendas in senate press releases. *Political Analysis*, *18*(1), 1–35. https://doi.org/10.1093/pan/mpp034

Grimmer, J., Roberts, M. E., & Stewart, B. M. (2021). Machine learning for social science: An agnostic approach. *Annual Review of Political Science*, *24*(1), 395–419. https://doi.org/10.1146/annurev-polisci-053119-015921

Grimmer, J., Roberts, M. E., & Stewart, B. (2022). *Text as data: A new framework for machine learning and the social sciences*. Princeton University Press.

Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, *21*(3), 267–297. https://doi.org/10.1093/pan/mps028

Grootendorst, M. (2022, March 11). *Bertopic: Neural topic modeling with a class-based TF-IDF procedure*. Retrieved November 7, 2022, from https://arxiv.org/abs/2203.05794

Guo, L. (2019). Media agenda diversity and intermedia agenda setting in a controlled media environment: A computational analysis of China's online news. *Journalism Studies*, *20*(16), 2460–2477. https://doi.org/10.1080/1461670X.2019.1601029

Gurajala, S., Dhaniyala, S., & Matthews, J. N. (2019). Understanding public response to air quality using tweet analysis. *Social Media and Society*, *5*(3), 205630511986765. https://doi.org/10.1177/2056305119867656

Habermas, J. (1991). *The structural transformation of the public sphere: An inquiry into a category of bourgeois society*. MIT press.

Hannigan, T., Haans, R. F. J., Vakili, K., Tchalian, H., Glaser, V. L., Wang, M., Kaplan, S., & Jennings, P. D. (2019). Topic modeling in management research. *The Academy of Management Annals*, *13*(2), 586–632. https://doi.org/10.5465/annals.2017.0099

Isoaho, K., Gritsenko, D., & Mäkelä, E. (2021). Topic modeling and text analysis for qualitative policy research. *Policy Studies Journal*, *49*(1), 300–324. https://doi.org/10.1111/psj.12343

Jacobi, C., Van Atteveldt, W., & Welbers, K. (2016). Quantitative analysis of large amounts of journalistic texts using topic modelling. *Digital Journalism*, *4*(1), 89–106. https://doi.org/10.1080/21670811.2015.1093271

Johnson, O., Hall-Phillips, A., Chung, T. L., & Cho, H. (2019). Are you connected through consumption? The role of hashtags in political consumption. *Social Media and Society*, *5*(4), 205630511988342. https://doi.org/10.1177/2056305119883427

Kaiser, J., Rauchfleisch, A., & Bourassa, N. (2020). Connecting the (far-)right dots: A topic modeling and hyperlink analysis of (far-)right media coverage during the US elections 2016. *Digital Journalism*, *8*(3), 422–441. https://doi.org/10.1080/21670811.2019.1682629

Kargar, S., & Rauchfleisch, A. (2019). State-aligned trolling in Iran and the double-edged affordances of Instagram. *New Media & Society*, *21*(7), 1506–1527. https://doi.org/10.1177/1461444818825133

King, G., Keohane, R. O., & Verba, S. (1994). *Designing social inquiry*. Princeton University Press.

Kligler-Vilenchik, N., Baden, C., & Yarchi, M. (2020). Interpretative polarization across platforms: How political disagreement develops over time on Facebook, Twitter, and WhatsApp. *Social Media and Society*, *6*(3). https://doi.org/10.1177/2056305120944393

Koltsova, O., & Koltcov, S. (2013). Mapping the public agenda with topic modeling: The case of the Russian LiveJournal. *Policy and Internet*, *5*(2), 207–227. https://doi.org/10.1002/1944-2866.POI331

Koltsova, O., & Shcherbak, A. (2015). 'LiveJournal Libra!': The political blogosphere and voting preferences in Russia in 2011–2012. *New Media & Society*, *17*(10), 1715–1732. https://doi.org/10.1177/1461444814531875

Koppers, L., Rieger, J., Boczek, K., & von Nordheim, G. (2021). Tosca: Tools for statistical content analysis. Retrieved January 15, 2013, from https://cran.r-project.org/web/packages/tosca/index.html

Krotz, F. (2007). The meta-process of mediatization as a conceptual frame. *Global Media and Communication*, *3*(3), 256–260. https://doi.org/10.1177/17427665070030030103. 030030103.

Kwon, K. H., Chadha, M., & Wang, F. (2019). Proximity and networked news public: Structural topic modeling of global Twitter conversations about the 2017 Quebec mosque shooting. *International Journal of Communication*, *13*(24), 2652–2675.

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A. L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D., & Van Alstyne, M. (2009). Computational social science. *Science*, *323*(5915), 721–723. https://doi.org/10.1126/science.1167742

Lind, F., Eberl, J. -M., Eisele, O., Heidenreich, T., Galyga, S., & Boomgaarden, H. G. (2021). Building the Bridge: Topic modeling for comparative research. *Communication Methods and Measures*, *16*(2), 1–19. https://doi.org/10.1080/19312458.2021.1965973

Lindgren, S. (2018). A ghost in the machine: Tracing the role of "the digital" in discursive processes of cybervictimisation. *Discourse & Communication*, *12*(5), 517–534. https://doi.org/10.1177/1750481318766936

Lu, Y., & Pan, J. (2020). Capturing clicks: How the Chinese government uses clickbait to compete for visibility. *Political Communication*, *38*(1–2), 1–32. https://doi.org/10.1080/10584609.2020.1765914

Maier, D., Baden, C., Stoltenberg, D., De Vries-Kedem, M., & Waldherr, A. (2022). Machine translation vs. multilingual dictionaries assessing two strategies for the topic modeling of multilingual text collections. *Communication Methods and Measures*, *16*(1), 19–38. https://doi.org/10.1080/19312458.2021.1955845

Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., Pfetsch, B., Heyer, G., Reber, U., Häussler, T., Schmid-Petri, H., & Adam, S. (2018). Applying LDA topic modeling in communication research: Toward a valid and reliable methodology. *Communication Methods and Measures*, *12*(2–3), 93–118. https://doi.org/10.1080/19312458.2018.1430754

Margolin, D. B. (2019). Computational contributions: A symbiotic approach to integrating big, observational data studies into the communication field. *Communication Methods and Measures*, *13*(4), 1–19. https://doi.org/10.1080/19312458.2019.1639144

McKernan, B. (2019). Digital texts and moral questions about immigration: Papers, please and the capacity for a video game to stimulate sociopolitical discussion. *Games and Culture*, *16*(4), 383–406. https://doi.org/10.1177/1555412019893882

Merrill, S., & Åkerlund, M. (2018). Standing up for Sweden? The racist discourses, architectures and affordances of an anti-immigration Facebook group. *Journal of Computer-Mediated Communication*, *23*(6), 332–353. https://doi.org/10.1093/jcmc/zmy018

Mohr, J. W., & Bogdanov, P. (2013). Introduction-Topic models: What they are and why they matter. *Poetics*, *41*(6), 545–569. https://doi.org/10.1016/j.poetic.2013.10.001

Möller, J., Trilling, D., Helberger, N., & van Es, B. (2018). Do not blame it on the algorithm: An empirical assessment of multiple recommender systems and their impact on content diversity. *Information, Communication & Society*, *21*(7), 959–977. https://doi.org/10.1080/1369118X.2018.1444076

Nelson, L. K. (2020). Computational grounded theory: A methodological framework. *Sociological Methods & Research*, *49*(1), 3–42. https://doi.org/10.1177/0049124117729703

Nelson, L. K., Burk, D., Knudsen, M., & McCall, L. (2018). The future of coding: A comparison of hand-coding and three types of computer-assisted text analysis methods. *Sociological Methods & Research*. https://doi.org/10.1177/0049124118769114

Ophir, Y., Walter, D., Arnon, D., Lokmanoglu, A., Tizzoni, M., Carota, J., D'Antiga, L., & Nicastro, E. (2021). The framing of COVID-19 in Italian media and its relationship with community mobility: A mixed-method approach. *Journal of Health Communication*, *26*(3), 161–173. https://doi.org/10.1080/10810730.2021.1899344

Pantti, M., Nelimarkka, M., Nikunen, K., & Titley, G. (2019). The meanings of racism: Public discourses about racism in Finnish news media and online discussion forums. *European Journal of Communication*, *34*(5), 503–519. https://doi.org/10.1177/0267323119874253

Papadouka, M. E., Evangelopoulos, N., & Ignatow, G. (2016). Agenda setting and active audiences in online coverage of human trafficking. *Information, Communication & Society*, *19*(5), 655–672. https://doi.org/10.1080/1369118X.2016.1139615

Patty, J. W., & Penn, E. M. (2014). Analyzing big data: Social choice and measurement. *PS - Political Science and Politics*, *48*(1), 95–101. https://doi.org/10.1017/S1049096514001814

Pöyhtäri, R., Nelimarkka, M., Nikunen, K., Ojala, M., Pantti, M., & Pääkkönen, J. (2021). Refugee debate and networked framing in the hybrid media environment. *International Communication Gazette*, *83*(1), 81–102. https://doi.org/10.1177/1748048519883520

Quinn, K. M., Monroe, B. L., Colaresi, M., Crespin, M. H., & Radev, D. R. (2010). How to analyze political attention with minimal assumptions and costs. *American Journal of Political Science*, *54*(1), 209–228. https://doi.org/10.1111/j.1540-5907.2009.00427.x

Reber, U. (2019). Overcoming language barriers: Assessing the potential of machine translation and topic modeling for the comparative analysis of multilingual text corpora. *Communication Methods and Measures*, *13*(2), 102–125. https://doi.org/10.1080/19312458.2018.1555798

Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using siamese BERT-networks. *EMNLP-IJCNLP 2019 - 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing, Proceedings of the Conference*, 3982–3992. https://doi.org/10.18653/v1/d19-1410

Rhidenour, K. B., Blackburn, K., Barrett, A. K., & Taylor, S. (2021). Mediating medical marijuana: Exploring how veterans discuss their stigmatized substance use on Reddit. *Health Communication*, *00*(00), 1–11. https://doi.org/10.1080/10410236.2021.1886411

Righi, R., Samoili, S., López Cobo, M., Vázquez-Prada Baillet, M., Cardona, M., & De Prato, G. (2020). The AI techno-economic complex system: Worldwide landscape, thematic subdomains and technological collaborations. *Telecommunications Policy*, *44*(6). https://doi.org/10.1016/j.telpol.2020.101943

Roberts, M. E., Stewart, B. M., & Airoldi, E. M. (2016). A model of text for experimentation in the social sciences. *Journal of the American Statistical Association*, *111*(515), 988–1003. https://doi.org/10.1080/01621459.2016.1141684

Robles, J. M., Velez, D., De Marco, S., Rodríguez, J. T., & Gomez, D. (2020). Affective homogeneity in the Spanish general election debate. A comparative analysis of social networks political agents. *Information, Communication & Society*, *23*(2), 216–233. https://doi.org/10.1080/1369118X.2018.1499792

Savolainen, L., Trilling, D., & Liotsiou, D. (2020). Delighting and detesting engagement: Emotional politics of junk news. *Social Media and Society*, *6*(4), 205630512097203. https://doi.org/10.1177/2056305120972037

Schroeder, R. (2018). *Social theory after the internet: Media, technology and globalization*. UCL Press.

Segal, Z. M., & Soffer, O. (2020). From weekly to daily: Computational analysis of periodical time cycles. *Journalism Studies*, *21*(14), 1952–1972. https://doi.org/10.1080/1461670X.2020.1807394

Shi, T., Kang, K., Choo, J., & Reddy, C. K. (2018). Short-text topic modeling via non-negative matrix factorization enriched with local word-context correlations. *The Web Conference 2018 - Proceedings of the World Wide Web Conference, WWW 2018*, 1105–1114. https://doi.org/10.1145/3178876.3186009

Simon, M., Welbers, K. C., Kroon, A., & Trilling, D. (2022). Linked in the dark: A network approach to understanding information flows within the Dutch Telegramsphere. *Information, Communication & Society*, 1–25. https://doi.org/10.1080/1369118X.2022.2133549

Slater, M. D., & Gleason, L. S. (2012). Contributing to theory and knowledge in quantitative communication science. *Communication Methods and Measures*, *6*(4), 215–236. https://doi.org/10.1080/19312458.2012.732626

Smith, N., & Graham, T. (2019). Mapping the anti-vaccination movement on Facebook. *Information, Communication & Society*, *22*(9), 1310–1327. https://doi.org/10.1080/1369118X.2017.1418406

Stier, S., Posch, L., Bleier, A., & Strohmaier, M. (2017). When populists become popular: Comparing Facebook use by the right-wing movement Pegida and German political parties. *Information, Communication & Society*, *20*(9), 1365–1388. https://doi.org/10.1080/1369118X.2017.1328519

Su, Y., Hu, J., & Lee, D. K. L. (2020). Delineating the transnational network agenda-setting model of mainstream newspapers and Twitter: A machine-learning approach. *Journalism Studies*, *21*(15), 2113–2134. https://doi.org/10.1080/1461670X.2020.1812421

Sweitzer, M. D., & Shulman, H. C. (2018). The effects of metacognition in survey research: Experimental, cross-sectional, and content-analytic evidence. *Public Opinion Quarterly*, *82*(4), 745–768. https://doi.org/10.1093/poq/nfy034

Teddlie, C., & Tashakkori, A. (2010). Overview of Contemporary Issues in Mixed Methods Research. *Sage Handbook of Mixed Methods in Social and Behavioral Research* (2nd ed., pp. 1–44). Thousand Oaks: SAGE Publications, Inc. https://dx.doi.org/10.4135/9781506335193

Thorson, K., Medeiros, M., Cotter, K., Chen, Y., Rodgers, K., Bae, A., & Baykaldi, S. (2020). Platform civics: Facebook in the local information infrastructure. *Digital Journalism*, *8*(10), 1231–1257. https://doi.org/10.1080/21670811.2020.1842776

Törnberg, A., & Törnberg, P. (2016). Combining CDA and topic modeling: Analyzing discursive connections between islamophobia and anti-feminism on an online forum. *Discourse & Society*, *27*(4), 401–422. https://doi.org/10.1177/0957926516634546

Tresch, A., & Feddersen, A. (2019). The (in)stability of voters' perceptions of competence and associative issue ownership: The role of media campaign coverage. *Political Communication*, *36*(3), 394–411. https://doi.org/10.1080/10584609.2018.1542416

van Atteveldt, W., & Peng, T. Q. (2018). When communication meets computation: Opportunities, challenges, and pitfalls in computational communication science. *Communication Methods and Measures*, *12*(2–3), 81–92. https://doi.org/10.1080/19312458.2018.1458084

van der Meer, T. G. L. A. (2018). Public frame building: The role of source usage in times of crisis. *Communication Research*, *45*(6), 956–981. https://doi.org/10.1177/0093650216644027

Van Der Velden, M., Schumacher, G., & Vis, B. (2018). Living in the past or living in the future? Analyzing parties' platform change in between elections, the Netherlands 1997–2014. *Political Communication*, *35*(3), 393–412. https://doi.org/10.1080/10584609.2017.1384771

Vayansky, I., & Kumar, S. A. P. (2020). A review of topic modeling methods. *Information Systems*, *94*, 101582. https://doi.org/10.1016/j.is.2020.101582

Von Nordheim, G., Boczek, K., & Koppers, L. (2018). Sourcing the sources: An analysis of the use of twitter and Facebook as a journalistic source over 10 years in the New York Times, the guardian, and Süddeutsche Zeitung. *Digital Journalism*, *6*(7), 807–828. https://doi.org/10.1080/21670811.2018.1490658

Vulić, I., De Smet, W., Tang, J., & Moens, M. F. (2015). Probabilistic topic modeling in multilingual settings: An overview of its methodology and applications. *Information Processing & Management*, *51*(1), 111–147. https://doi.org/10.1016/j.ipm.2014.08.003

Wahlström, M., Törnberg, A., & Ekbrand, H. (2020). Dynamics of violent and dehumanizing rhetoric in far-right social media. *New Media & Society*, *23*(11), 146144482095279. https://doi.org/10.1177/1461444820952795

Waldherr, A., Geise, S., Mahrt, M., Katzenbach, C., & Nuernbergk, C. (2021). Toward a stronger theoretical grounding of computational communication science. *Computational Communication Research*, *3*(2), 1–28. https://doi.org/10.5117/ccr2021.02.002.wald

Walter, D., & Ophir, Y. (2020). Strategy framing in news coverage and electoral success: An analysis of topic model networks approach. *Political Communication*, *38*(6), 1–24. https://doi.org/10.1080/10584609.2020.1858379

Xu, K., Liu, F., Mou, Y., Wu, Y., Zeng, J., & Schäfer, M. S. (2020). Using machine learning to learn machines: A cross-cultural study of users' responses to machine-generated artworks. *Journal of Broadcasting & Electronic Media*, *64*(4), 566–591. https://doi.org/10.1080/08838151.2020.1835136

Yang, A., Choi, I. M., Abeliuk, A., & Saffer, A. (2021). The influence of interdependence in networked publics spheres: How community-level interactions affect the evolution of topics in online discourse. *Journal of Computer-Mediated Communication*, *26*(3), 148–166. https://doi.org/10.1093/jcmc/zmab002

Yan, X., Guo, J., Lan, Y., & Cheng, X. (2013). A biterm topic model for short texts. In *Proceedings of the 22nd international conference on World Wide Web*, Rio de Janeiro, Brazil, (pp. 1445–1456).

Yan, P., Schroeder, R., & Stier, S. (2021). Is there a link between climate change scepticism and populism? An analysis of web tracking and survey data from Europe and the US. *Information, Communication & Society*, *0*(0), 1–40. https://doi.org/10.1080/1369118X.2020.1864005

Yarchi, M., Baden, C., & Kligler-Vilenchik, N. (2020). Political polarization on the digital sphere: A cross-platform, over-time analysis of interactional, positional, and affective polarization on social media. *Political Communication*, *38*(1–2), 1–42. https://doi.org/10.1080/10584609.2020.1785067

Ying, L., Montgomery, J. M., & Stewart, B. M. (2021). Topics, concepts, and measurement: A crowdsourced procedure for validating topics as measures. *Political Analysis*, *30*(4), 570–589. https://doi.org/10.1017/pan.2021.33

Zeng, J., Chan, C. H., & Fu, K. W. (2017). How social media construct "truth" around crisis events: Weibo's rumor management strategies after the 2015 Tianjin blasts. *Policy and Internet*, *9*(3), 297–320. https://doi.org/10.1002/poi3.155

Zeng, J., Chan, C., & Schäfer, M. S. (2020). Contested Chinese dreams of AI? Public discourse about artificial intelligence on WeChat and People's daily online. *Information, Communication & Society*, *25*(3), 319–340. https://doi.org/10.1080/1369118X.2020.1776372

Zhang, X. (2018). Visualization, technologies, or the public? Exploring the articulation of data-driven journalism in the Twittersphere. *Digital Journalism*, *6*(6), 737–758. https://doi.org/10.1080/21670811.2017.1340094

Zhao, Q., Qin, Z., & Wan, T. (2011). Topic modeling of Chinese language using character-word relations. In *International Conference on Neural Information Processing* (pp. 139–147). Springer, Berlin, Heidelberg.

Zhao, X., Zhan, M., & Jie, C. (2018). Examining multiplicity and dynamics of publics' crisis narratives with large-scale twitter data. *Public Relations Review*, *44*(4), 619–632. https://doi.org/10.1016/j.pubrev.2018.07.004

Zheng, P., & Shahin, S. (2020). Live tweeting live debates: How twitter reflects and refracts the US political climate in a campaign season. *Information, Communication & Society*, *23*(3), 337–357. https://doi.org/10.1080/1369118X.2018.1503697