A Survey of Computational Framing Analysis Approaches

Mohammad Ali
College of Information Studies
University of Maryland, College Park
mali24@umd.edu

Naeemul Hassan
Philip Merrill College of Journalism
College of Information Studies
University of Maryland, College Park
nhassan@umd.edu

Abstract

Framing analysis is predominantly qualitative and quantitative, examining a small dataset with manual coding. Easy access to digital data in the last two decades prompts scholars in both computation and social sciences to utilize various computational methods to explore frames in large-scale datasets. The growing scholarship, however, lacks a comprehensive understanding and resources of computational framing analysis methods. Aiming to address the gap, this article surveys existing computational framing analysis approaches and puts them together. The research is expected to help scholars and journalists gain a deeper understanding of how frames are being explored computationally, better equip them to analyze frames in large-scale datasets, and, finally, work on advancing methodological approaches.

1 Introduction

Vaccine hesitancy has long been recognized as a problem despite research evidence favoring the vaccine's effectiveness (Sallam, 2021). Understanding how the vaccination is framed by news media might provide a solution to vaccine hesitancy because a frame determines "how [people] evaluate [a problem] and choose to act upon it" (Entman, 1993, p. 54). Like this, exploration of many other problems (e.g., gun violence) warrants analysis of frames, especially in large-scale datasets in this era.

Traditionally, researchers explore frames using qualitative and quantitative methods that require manual labor and can handle small amounts of data (D'angelo, 2018; Reese et al., 2001). Production of and easy access to large volumes of digital data in the last two decades prompt scholars to harness the exploration of frames in such big data computationally (Card et al., 2015; Liu et al., 2019; Walter and Ophir, 2019; van Atteveldt and Peng, 2018).

Prior studies proposed various computational methods (e.g., topic modeling and neural network). As the scholarship is growing, a scarcity appeared regarding a comprehensive understanding and resources of computational framing analysis methods (Nicholls and Culpepper, 2021; Sanfilippo et al., 2008). Researchers might be confused with multiple approaches to this analysis, raising questions: how many computational framing analysis methods exist, and which one they should apply?

To address the problem and help researchers with such questions, we survey existing computational framing analysis approaches and put the methods and relevant resources together. As such, the survey is guided by the following three research questions:

RQ1. What computational methods do researchers use to explore frames in large-scale datasets?

RQ2. How do researchers conceptualize a frame in computational framing analysis studies?

RQ3. How do researchers use computational methods in exploring frames?

The primary contributions of this article are: a) it provides a comprehensive understanding and resources of existing computational framing analysis methods and puts them together for interested scholars to gain deeper knowledge and start building on that, and b) it adds new thoughts to the ongoing discussion on advancing the computational methods of framing analysis.

2 What is Frame or Framing?

This section provides a conceptual understanding of framing. A classic example of framing concerns a debate over whether to permit *Ku Klux Klan* to hold a public rally. One news story with the headline "Ku Klux Klan Tests OSU's Commitment to Free Speech" reported the rally as a free speech issue, while another one with the headline "Possible Ku Klux Klan Rally Raises Safety Concerns" reported it as a disruption of public order. As reflected in the headlines, the two stories used different frames. People who read the free speech news story expressed higher tolerance toward KKK's

Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, pages 9335 - 9348
December 7-11, 2022

Figure 1: Framing devices deployed in the headlines of two news reports published by *The New York Times* and *The Guardian* on the 2022 Buffalo mass shooting.

rally compared to those who read the public order news story (Nelson et al., 1997, p. 581). Figure 1 shows similar frames deployed in two news headlines on the 2022 Buffalo mass shooting.

Scholars are not agreed upon any unified framing definition (Hertog and McLeod, 2001; Van Dijk, 2016). However, a prominent definition, widely used in both traditional and computational framing studies, was provided by Entman (1993). He says:

> To frame is to *select some aspects of a perceived reality and make them more salient in a communicating text, in such a way as to promote a particular problem definition, causal interpretation, moral evaluation, and/or treatment recommendation* for the item described. (p. 52)

As per this definition, a frame is largely determined by its outcome effects, such as four functions: a) defining problems, b) diagnosing causes, c) making judgments, and d) suggesting remedies. The functions depend on how some selected aspects of "perceived" reality are made salient. In 2003, he defined it a bit differently, "Framing entails *selecting and highlighting some facets of events or issues, and making connections among them so as to promote a particular interpretation, evaluation, and/or solution"* (Entman, 2003, p. 417). This definition seems to have made a few shifts, such as from "causal interpretation" to "interpretation," from "moral evaluation" to "evaluation," and from "treatment recommendation" to "solution." The salient aspects are also interconnected.

While approaching frames as cultural phenomena, Hertog and McLeod (2001) identified a frame as a cultural "[structure] of meaning that includes a set of core concepts and ideas," including "conflicts, metaphors, myths, and narratives" (p. 160). A frame has also been explained as "a central organizing idea... for making sense of relevant events, suggesting what is at issue" (Gamson and Modigliani, 1989, p. 3). Reese et al. (2001) defined a frame from the sociological perspective and focused on six aspects (italicize): "Frames are *organizing principles* that are socially *shared* and *persistent* over time, that work *symbolically* to meaningfully *structure* the social world" (p. 11). In a recent definition, D'angelo (2018) defined news framing as "how journalists, their sources, and audiences work within conditions that shape the messages they construct as well as the ways they understand and interpret these messages" (p. xxiv).

To describe a frame's aspect highlighting some selected facets of an issue or event, Fairhurst (2005) utilized an analogy that "choosing language to frame people's actions and events is like moving a telescope into position" (p. 125). The selected aspects are then coherently organized in a way to make an argument, which finally promotes a particular interpretation, evaluation, and solution. This organization of selected aspects could even be subtle, as framing also "refers to subtle alterations in the statement or presentation of judgment and choice problems" (Iyengar, 1994, p. 11). Another crucial aspect of framing is "to choose one particular meaning (or set of meanings) over another" (Fairhurst and Sarr, 1996, p. 3) that is also supported by Entman (1993), who says a frame "operates by selecting and highlighting some features of reality while omitting others" (p. 53).

Contexts in Framing. A frame is considered context-sensitive. It is shaped in four locations: i) communicator, ii) texts, iii) receiver, and iv) culture (Entman, 1993). The culture is the stock of commonly invoked frames and explained as (a part of) contexts. A news report's content is fully comprehensible when its contextual information is at the disposal of readers. They interpret a frame and its meaning following contextual information (Baden and D'Angelo, 2018; Tewksbury and Riles, 2018).

Framing Devices. Framing devices can be defined as tools that are used to make a piece of information more salient, which is, in other words, "making a piece of information more noticeable, meaningful, or memorable to audiences" (Entman, 1993, p. 53). While conceptualizing a frame, we accumulated framing devices (see Table 1). To make the list concise and convenient, we combined similar devices and put them into four groups: a) content, b) action, c) context, and d) communicator. The devices or tools can be used to provide either higher or lower salience to selected aspects of reality. In some cases, multiple devices can be applied together as a new device. For example, jargon, metaphors, and contrast can together be used to develop a "story" (Fairhurst and Sarr, 1996).

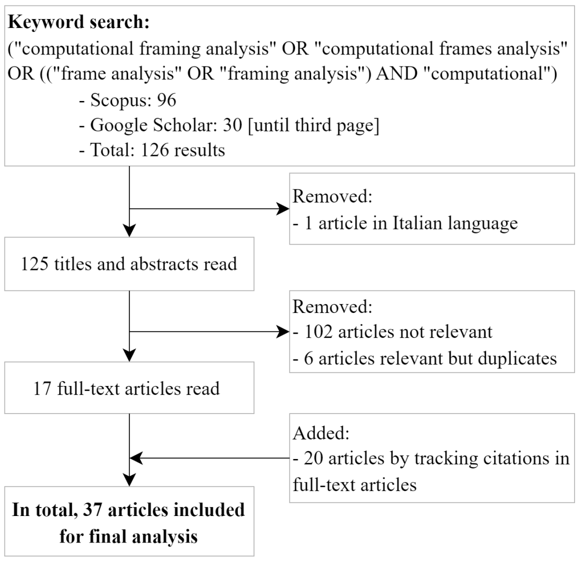

Figure 2: Summary of the Paper Selection Method

3 Method

We utilized three ways to identify and select relevant articles for a comprehensive understanding of computational framing analysis methods. First, we searched on Scopus, an abstract and citation database of Elsevier, using relevant keywords: ("computational framing analysis" OR "computational frames analysis" OR (("frame analysis" OR "framing analysis") AND "computational"). It provides 95 articles in the English language. We manually read their abstracts and sorted out 13 articles relating to computational framing analysis. In the sorting process, we read the articles' method sections if needed to make the decision. Other 82 articles were excluded due to their irrelevance. The excluded articles were related to "frames" in other fields, such as building structures (e.g., 2D plane frames) and mechanical engineering. Second, we searched on Google Scholar using the exact keywords and included articles until the third page as no relevant article was found on the third page. This gave us ten relevant articles. Six articles were common in both the Scopus and the Google Scholar searches, resulting in 17 unique articles from both sources. Third, while reading through the 17 selected articles, we tracked down 20 more relevant articles cited in some of those articles. The 20 articles did not appear in the Scopus and Google Scholar searches probably because of the different keywords and phrases used in their titles and abstracts.

Finally, we got a total of 37 articles selected for this survey (see Figure 2). The articles involve journals and conferences in both computation and social science disciplines. Reading through the articles and their supplemental materials (e.g., coding schema guiding the annotation), if any, we utilized an inductive way to scrutinize various aspects, including a) framing conceptualization, b) functions of computational framing analysis approaches, and c) results and their interpretation. We reported available datasets, codes, and other relevant resources, if any.

4 Analysis

This section presents an analysis of the selected articles in two broad parts. The first part answers RQ1, and the second part answers RQ2 and RQ3. Table 2 summarizes the articles, identified approaches, codebook, corpora, domains, and resources.

Codebook, Corpora, & Approaches (RQ1). Analysis of the articles identified at least nine approaches and three major coding schema and annotated corpora for computational framing analysis. The approaches are in the categories of supervised, unsupervised, and mixed methods. A supervised method usually needs an annotated subset of data. Here, the model is first trained on a labeled dataset (training data) and then applied in a new similar dataset (test data) to classify or predict each instance (Kotsiantis et al., 2007). In contrast, an unsupervised method does not need any pre-annotated datasets. Instead, it explores all unlabeled data.

Conceptualization & Functions (RQ2 & RQ3). As a way of answering RQ2 and RQ3, we explore how researchers conceptualize frames and utilize computational methods in analyzing frames in each approach, codebook, and corpora.

4.1 Codebook & Corpora

4.1.1 Policy Frames Codebook

Boydstun et al. (2013) and Boydstun et al. (2014) proposed a codebook named "policy frames codebook" (PFC). The PFC consists of 14 categories of "frame dimensions" and an "other" category. The dimensions include "economic frames," "capacity and resources frames," "morality frames," etc. For example, a news report is labeled as an economic frame if it focuses on "the costs, benefits, or monetary/financial implications of the issue (to an individual, family, community, or to the economy as a whole)" (Boydstun et al., 2014, p. 6).

They developed the codebook through brainstorming and iteration of applying it to random texts. With the codebook, they deployed 3,033 coders to manually code three sets of articles on immigration, tobacco, and same-sex marriage. Using the labeled documents, they finally developed a logistic regression binary text classifiers (i.e., present or absent) (Boydstun et al., 2013, 2014).

4.1.2 Media Frames Corpus

Using PFC, Card et al. (2015) offered a manually-annotated corpus of news reports named "media frames corpus" (MFC). The news reports were collected from three domains: immigration, smoking, and same-sex marriage. The MFC was applied in other studies (e.g., Field et al., 2018). Card et al. (2015) annotated the three datasets based on PFC's 15 framing dimensions (Boydstun et al., 2013). The authors, however, did not apply the annotations to any new datasets. In 2016, they added four more categories—pro, neutral, anti, and irrelevant.

Conceptualization in PFC & MFC. Boydstun et al. (2013, 2014) conceptualized framing by resorting to the widely used framing definition of Entman (1993). Overall, they put "language" at the center of identifying and analyzing frames. PFC's development is motivated by three framing concepts: a) frame selection varies based on various situations, b) frames evolve over time, and c) frames spread across issues, geographic locations, and institutions or organizations. Card et al. (2015) also used Entman (1993)'s definition in conceptualizing frames. They focused on some framing elements that work coherently as a framing package.

Review. The authors conceptualized frames with existing framing definitions. However, framing aspects they mentioned (e.g., Entman, 1993) were not utilized in developing the 15 "framing dimensions." Considering the development process and broader definitions of each frame, the 15 dimensions seem to be more fit with "topics," not frames. As per the framing theory, the categorization of these dimensions looks arbitrary and too broad to understand a frame's nuances. For example, a text is identified as an "economic frame" if it focuses on anything of the whole economy. Let's consider the *Ku Klux Klan's* example mentioned above. As per MFC's 15 dimensions, both KKK news reports could probably be identified as a "law and order, crime and justice frame" under the PFC. Here, it does not answer the "how" question at all. The dimensions, however, can be considered as topics. The MFC corpus inherited the same limitations as it was developed using the PFC codebook.

4.1.3 Gun Violence Frame Corpus (GVFC)

This article identified another annotated corpus named "Gun Violence Frame Corpus" (GVFC). It was applied in neural network-based models discussed later. In this dataset, the authors manually annotated 1,300 news headlines collected from 21 U.S. news media outlets. Using nine pre-defined codes drawn from literature, multiple coders annotated the headlines. Finally, they used a BERT model and made a frame prediction classifier. Its overall accuracy is 84.23

Conceptualization. Liu et al. (2019) used Entman (1993)'s prominent definition to conceptualize framing. They highlighted various ways of constructing frames, such as word choice and labeling by journalists "to promote a certain side" (p. 504). The authors also focused on generic versus issue-specific frames. In terms of manual codes, they applied a deductive approach— first defining some frames and then manually labeling news articles into those pre-defined frames.

Review. The article briefly conceptualized a frame and included the aspects of widely-used framing definition (e.g., Entman, 1993). However, all the framing codes in GVFC were not defined following how the framing was conceptualized. For example, a code was defined in the category of politics "... as long as [a] news headline mentions a politician's name" which seems not aligned with the nuances of their conceptualizations.

4.2 Computational Approaches

4.2.1 Topic Modeling

Various prior studies utilized topic modeling (TM) to explore frames (e.g. DiMaggio et al., 2013).

Method. The TM algorithm discovers latent themes in a large collection of documents (Blei, 2012). A topic is a probability distribution over a fixed vocabulary (p. 78). The algorithm produces a number (*k*) of lists of words based on the words' higher probability of being in a list. Each list of words is considered to be a topic, and each topic has a different probability distribution. The latent Dirichlet allocation (LDA) topic model provides an assignment of each document to the topic(s). As a mixed-membership model, each of its documents may be assigned to multiple topics, considering that a document could have elements of multiple topics. DiMaggio et al. (2013) used the LDA topic modeling to explore frames. They view each topic as a frame, saying that a topic "includes terms that call attention to particular ways" (p. 593).

Conceptualization. In the study of DiMaggio et al. (2013), they conceptualized a frame as "a set of discursive cues (e.g., words, images, and narrative) that suggests a particular interpretation of a person, event, organization, practice, condition, or situation" (p. 593). They cited Gamson et al. (1992)'s definition that a frame is "a central organizing principle that holds together and gives coherence and meaning to a diverse array of symbols." They considered each topic as a frame.

Review. Here, the conceptualization of a frame looks consistent with the overall framing idea. However, the topic model's output (i.e., lists of words) and their interpretation seem not aligned with framing aspects. A list of words in the topic model comes without any connection among them due to its features (e.g., bag-of-words). The interpretation of each word list in DiMaggio et al. (2013) also indicates it as a theme or issue, not a frame. For example, they reported the results by utilizing words like "highlight," "emphasize," and "concerned with" (e.g., this topic *highlights* legislative actions). Framing nuances like a problem and causal interpretation could not be extracted here.

4.2.2 Structural Topic Modeling (STM)

Method. The STM model was also used to explore frames (e.g., Roberts et al., 2014). Compared to LDA topic modeling (Blei, 2012), STM allows including metadata or covariates in the model. With metadata (e.g., political ideology and time) added to the dataset and model, the STM allows researchers to interpret how the topics are associated with those metadata. For example, in terms of political ideology, such as conservatives and liberals, researchers might identify a topic as more aligned with conservatives and another topic with liberals. Metadata can also be used in predicting the topics' prevalence by metadata (Gilardi et al., 2021; Nicholls and Culpepper, 2021).

In their study exploring topics in a corpus of newspaper texts, Gilardi et al. (2021) used some covariates, including time. Their results show how the topics are distributed over time across various states in the U.S. Since the authors followed DiMaggio et al. (2013)'s argument of considering a topic as a frame, their results' interpretation also focuses on themes or topics, instead of frames.

Conceptualization. Gilardi et al. (2021) conceptualized a frame with Gamson et al. (1992) definition that a frame can be understood as a "storyline or unfolding narrative about an issue" (p. 385). In terms of exploring frames by STM, Gilardi et al. (2021) relied on DiMaggio et al. (2013) argument that topics identified through TM can be viewed as frames.

Review. Like the topic modeling approach (Gilardi et al., 2021), the STM algorithm is also constrained by considering a topic as a frame. So, the STM contains similar limitations in terms of framing analysis. Compared to topic modeling, the STM offers additional insights into the topics or themes through the analysis of covariates. Both methods are based on the bag-of-words idea, indicating the lack of semantic contextualization needed for exploring frames.

4.2.3 Hierarchical Topic Modeling

Method. Studies also used hierarchical topic modeling (HTM) to explore frames. Nguyen (2015) and Nguyen et al. (2015) introduced an HTM model named "Supervised Hierarchical Latent Dirichlet Allocation (SHLDA)" that aims to analyze frames in a large dataset. As the SHLDA works, each document in the corpus is associated with a continuous level of scores (e.g., conservative vs. liberal ideology). It produces a hierarchy of topics, where the first-level nodes are considered agendas and the second-level nodes as frames. Documents' scores help explain how the topics are framed concerning respective people's positions. Its document generative process combines the hierarchical LDA and hierarchical Dirichlet process (HDP). The authors applied it to three datasets and conducted qualitative and quantitative analyses to validate the models' agenda and frames.

Conceptualization. Nguyen (2015) also used the framing definition of Entman (1993) in conceptualizing a frame. However, unlike Gilardi et al. (2021), Nguyen (2015) considered a topic as an agenda (e.g., what topics are talked about) and a sub-topic as a second-level agenda or a frame (e.g., how these topics are talked about).

Review. As elaborated above, the SHLDA is one step ahead of topic modeling. However, a crucial incongruity remains in how they conceptualized a frame (e.g., sub-topics) and interpreted the results. Though there is a lack of unified framing definition, the idea of considering a sub-topic as a frame does not align with traditional framing conceptualization (Entman, 1993; McCombs et al., 1997; Ghanem, 1997). Like many prior framing studies, the SHLDA output might also be considered as simply topics and their relevant attributes, not frames. Moreover, Nguyen (2015)'s qualitative analysis to validate the output as frames is not systematically executed, and the presentation of its results does not illustrate any framing aspects (Entman, 1993)

4.2.4 Cluster Analysis

Method. The k-means clustering algorithm is another unsupervised approach used to explore frames. Burscher et al. (2016) conducted two k-means clustering in a dataset. One includes all words, and another includes selected words (i.e., nouns, adjectives, and adverbs). After creating document vectors with TFIDF in both groups, they conducted k-means clustering to find clusters. As a centroid-based clustering approach works, a certain number of clusters (k) is specified in advance, and each cluster is represented by its center. They select the number of clusters (k) using the "elbow method." Each document is assigned to a cluster based on its relatively closer distance to that cluster center (Burscher et al., 2016). Unlike topic modeling, k-means clustering is a single-membership approach where each document generally belongs to one cluster.

Conceptualization. Burscher et al. (2016) conceptualized a frame in terms of "word frequencies" and mentioned words as highly reliable and less biased in producing frames. They "used word frequencies as features [of a frame] in [their] cluster analyses" (p. 533). They utilized traditional framing definition partially (e.g., presence or absence of certain keywords, stock phrases) (Entman, 1993).

Review. As Burscher et al. (2016) conceptualized and interpreted frames in terms of word frequencies and co-occurrences, the framing devices listed in Table 1 suggest that word(s) are simply one of the many devices to construct a frame. They utilized such conceptualization that does not help explore frames despite their acknowledgment that "based on plain word features, a cluster analysis cannot reveal complex semantic and logical relationships like causality" (Burscher et al., 2016, p. 541). As a single-membership approach, this method is also against one of the core framing ideas that a framing device may belong to multiple frames. The results were presented with words, including "refer to." For example, "cluster B5 *refers to* nuclear power ... in Iran" (p. 439). The results indicate these as a topic or issue. It does not indicate "how" the "nuclear" issue was discussed and evaluated as a problem. Both conceptualization and output seem to illustrate certain topics, not frames.

4.2.5 Neural Network Model

Method. Some studies utilized the neural network approach to build frame-identifying classifiers and analyzed frames in various text documents (e.g., news reports and tweets). Mainly, two annotated datasets namely, MFC and GVFC, were used in building these models.

MFC was utilized in a number of such studies, including probabilistic soft logic (PSL) (Johnson et al., 2017), LSTM neural network (Naderi and Hirst, 2017), recursive neural network (Ji and Smith, 2017), and transformer-based language models such as BERT and RoBERTa (Khanehzar et al., 2019; Cabot et al., 2020; Mendelsohn et al., 2021). Some studies used MFC's annotated news reports partially and some used the full corpus.

Manually annotating the GVFC dataset, Liu et al. (2019) used it to build a classifier using BERT. It was later applied in other studies (e.g., Akyürek et al., 2020; Tourni et al., 2021; Bhatia et al., 2021).

Conceptualization. As mentioned above, Liu et al. (2019) used traditional framing definitions (e.g., Entman, 1993) while conceptualizing a frame. The studies applying MFC in building a neural network-based classifier also conceptualize it by drawing works from prior studies in both social and computational science.

Review. In terms of the approach, both groups of studies seem to have applied the state-of-the-art pre-trained models based on transfer learning that looks promising for advancing computational framing analysis. However, the quality of the annotated training dataset appears not up to the mark, which is reflected in the lack of results interpretation in those studies. As reviewed above, the MFC dataset seems more about categorizing a text into broad topics (e.g., "economic frames"), not frames. The subsequent studies applying MFC dataset also did not adequately justify MFC's 15 dimensions as frames. Their results mainly focused on the accuracy of the model built on MFC training dataset, but not whether the results provide framing nuances.

Compared to MFC, GVFC's annotations look more coherent but still lack in capturing framing nuances, as mentioned above in sub-section 4.1.3. For example, based on GVFC's "politics" code, Liu et al. (2019) interpreted its result saying, "it appears that news media of all types have largely *politicized* the gun violence issue right after each major mass shooting" (p. 511). Here, the *politicization* result and its interpretation do not align with how the code is defined. The results might indicate the texts "discussed" "a politician" or politics, which is a simple topic or an issue, not any major framing element like problem definition and its coherent argument.

4.2.6 Parsing Semantic Relations

Another line of computational framing analysis relates to the exploration of semantic relations, going beyond the bag-of-word model.

Method. Sturdza et al. (2018) operationalized Entman (1993)'s four framing elements as their semantic relations in texts. This approach proposed the utilization of a rule-based system that uses existing computational software, such as TurboParser, and implicature rules. Using the parser, the author proposed identifying syntactic structures in texts and then using a set of rules to transform the syntactic structure into semantic networks. The networks determine the semantic roles of each word (e.g., actors, events) through a set of sentiment analysis implicature rules using a sentiment lexicon.

On the other hand, Ziems and Yang (2021) computationally parsed various attributes (e.g., race) of police shooting victims in news reports and explored how differently they are portrayed in news media. They called it "entity-centric framing." A recent study by Yu (2022) looked at iterative adverbs (e.g., again) in the political discourse considering the adverbs evoke different attitudinal subtexts. After extracting sentences with relevant adverbs, the author grouped the sentences through *k*-means clustering and identified the most representative keywords in each cluster by a keyword mining tool.

Conceptualization. In conceptualizing a frame, Sturdza et al. (2018) relied on four framing elements of Entman (1993, p. 52). However, two other studies lack adequate conceptualization of framing. For instance, Ziems and Yang (2021) mainly explored "entity-centric" frames but did not elaborate on it from existing literature.

Review. Compared to the topic modeling method, this approach looks innovative in terms of understanding semantic relations between words and phrases. However, the idea seems not adequately exploited in understanding the nuances of frames. For example, Sturdza et al. (2018) did not apply the operationalization in a practical dataset. Results of Ziems and Yang (2021) reported frequency and correlations while Yu (2022)'s results ended up with clustering and keywords, instead of exploring the coherent argument and relations among various framing devices. However, by its design, the semantic relations approach holds the potential for being used in advancing the computational methods of framing analysis.

4.2.7 Frequency-based Model

Method. This model proposed using QDA Miner and its affiliated WordStat program to extract words, and phrases, and examine their repetitions across the corpus (Kang and Yang, 2022). In this model, Sanderink (2020) proposed little changes, which is to first determine certain frames (e.g., energy security) by reviewing prior scholarship. Researchers then prepare a codebook using QDA Miner. The codebook comprises words, phrases, and rules that capture various elements relating to each of the pre-determined frames. Finally, WordStat was used to calculate the frequency of words and phrases relating to each frame.

Conceptualization. Scholars in this approach defined a frame in terms of word recurrence in a document. They also highlighted the ways of editing, interpreting, organizing, and presenting information for particular news content to be framed. They compared a frame with a theme.

Review. The frame was not appropriately conceptualized here, as per the existing framing definitions (e.g., Entman, 1993). The consideration of only the frequency of words does not compromise the coherent meanings of frames.

4.2.8 FrameAxis

Method. FramAxis model explores "microframes," which is operationalized as a pair of antonyms, such as legal versus illegal and fast versus slow.

The antonyms are obtained from WordNet. Then, the authors compute the bias of each microframe (average contribution of all words in a document to the microframe) and the intensity of each microframe (how strongly it is presented in documents). The microframes are analyzed along with the agent-object-action patterns identified by the semantic role labeling (SRL) model in the corpus.

Conceptualization. A frame in this approach was conceptualized utilizing features of existing definitions. For example, they highlighted presenting some selected aspects of an issue and making them more salient, which aims to promote certain values, interpretations, or solutions.

Review. Though the framing conceptualization is derived from prominent framing definitions, the core aspect of FrameAxis is the pair of antonyms, which again limits the coherent argument, problem definition, and other framing elements.

4.2.9 Analysis of Topic Model Networks

Walter and Ophir (2019) proposed this mixed method approach, "Analysis of Topic Model Networks" (ANTMN), that combines topic modeling and semantic network analysis. It was applied in other studies (e.g., Ophir et al., 2021).

Method. ANTMN includes three steps. First, the authors apply LDA topic modeling (Blei, 2012) to the dataset. They label each topic by qualitatively examining three types of information: a) words with the highest loading over each topic, b) prevalent and exclusive words in each topic, and c) full documents that are the most representative of each topic. Second, ANTMN creates a semantic network, where the topics serve as nodes, and topics' similarity relationships serve as edges. The relationship is calculated based on the topics' co-occurrence in the documents. The output provides a fully connected, undirected, and weighted network. Finally, a community detection algorithm was used to cluster the topics into various communities in the network based on the topics' prevalence in similar documents (Walter and Ophir, 2019).

Conceptualization. As the authors noted, ANTMN can analyze emphasis frames (e.g., highlighting one side), not equivalency frames (e.g., gain vs. loss issue). They conceptualize a frame as "a communit[y] in a network of topics" (p. 248), based on linguistic patterns. Borrowing van Atteveldt and Peng (2018)'s idea of arranging various framing devices around an overarching idea (e.g., a cluster of relevant framing devices), they consider each topic in topic modeling as a framing device. The cluster of topics was named as a frame in ANTMN. They embraced the patterns of a frame that "*repeatedly invokes the same objects and traits, using identical or synonymous words and symbols in a series of similar communications that are concentrated in time*" (Entman et al., 2009, p. 177).

Review. A few things seem to have restricted ANTMN as a framing analysis model. As per the framing conceptualization, the topics (aka framing devices) under each network community need to be coherently connected with each other to render a coherent framing argument. The authors did not explain how the devices are coherently interconnected. This lack is reflected in the interpretation of the results. For instance, they reported a framing result, saying that "the largest community on the right *consisted of* topics about the cultural and economic consequences.... Articles dominated by these topics *portrayed* the impact of diseases on the economy at large.... (Walter and Ophir, 2019, p. 259). Here, the authors mentioned topics' names and what these topics portray with words like *"consists of"* and *"portrayed."* The results did not provide a coherent argument of the problem or how one aspect is interconnected with another. Though the output demonstrated some topics, the authors' claim of the communities as frames is not supported with adequate evidence.

Despite the authors' claim of this method as unsupervised, manual human labor is still needed in at least two places: a) an examination of words and documents to label topics and b) an interpretation of findings. However, no systematic method was provided for executing the manual analysis.

5 Discussion and Conclusion

In this article, we surveyed 37 empirical studies and reported on nine computational approaches, three coding schema, and annotated corpora of how they conceptualize frames and utilize various computational methods to explore frames in large-scale datasets. Overall, existing methods and relevant resources are put together in this article. In the absence of a comprehensive understanding and resources of computational framing analysis methods, this article's insights will benefit framing scholars, especially those who are new, to gain deeper knowledge in this single article and build on that in further exploring frames in big data.

Algorithmic Functions. As demonstrated above, most algorithms used in computational framing analysis were not originally built for this purpose. For example, LDA topic modeling is basically built to find broader themes in a large corpus (Blei, 2012). The works of Liu et al. (2019) and Walter and Ophir (2019), however, seem to be innovative in terms of their efforts to build a new or modified method to explore comparatively more nuances of frames (Nicholls and Culpepper, 2021). As state-of-the-art models, neural networks appeared promising, but appropriate training datasets need to be developed and used for that.

Conceptualization of Frames. Though the computational methods mostly conceptualized a frame with prominent definitions (e.g., the definition of Entman, 1993), some of the methods embraced framing aspects partially. Some studies ended up operationalizing a frame in a way that is not supported by the core framing aspects. For instance, Boydstun et al. (2013, 2014) include its main aspects in developing PFC, which defined the 15 dimensions as "topics" in the name of frames. Nguyen (2015) simply equated a frame with second-level agendas or sub-topics without adequate conceptual support. Though Liu et al. (2019) and Walter and Ophir (2019) provided relatively stronger conceptualization, their results suggest that Liu et al. (2019)'s coding schema and Walter and Ophir (2019)'s network communities still lack in providing coherent definition and causal interpretation arguments.

Interpretation of Results. Even if some studies conceptualized frames in a relatively comprehensive way, their results presentation and interpretation rarely went above describing relevant topics and themes, not frames, as their results lack illustrating the coherent problem, causal evaluations, or potential recommendations. An example mentioned under ANTMN above demonstrated such evidence. Similar gaps in terms of framing conceptualization and presentation of results and interpretations remain in other approaches as well (e.g., topic modeling and cluster analysis).

Use of Framing Devices. The bag-of-word approach automatically excludes from analysis many potential framing devices listed in Table 1. The approaches examined in this article mostly utilize only one framing device (i.e., words). Considering the fact that framing analysis is a comprehensive approach involving multiple theoretical and practical aspects (D'angelo, 2018; Golan, 2021), even the qualitative framing analysis through manual labor is challenging work. From that perspective, computational approaches are in the nascent stage in addressing this social science problem of framing analysis. So, the scholarship needs better computational methods and tools that might explore frames as close as possible. For example, computational approaches might want to retrieve the problem definition and causal interpretation by including more framing devices (see Table 1) by going beyond the analysis of "words" in future studies.

Overall, this survey article contributed to the literature on computational framing analysis in several ways. As the first survey paper, it put together existing computational framing analysis methods and resources in one place, which can benefit future scholars as at least a source of gaining more comprehensive knowledge on computational framing analysis approaches. With this knowledge, they can start further exploring frames in big data and advancing computational framing analysis methods. This article also contributed to the ongoing discussion and scholarly efforts on further improving the computational tools in framing analysis.

Open Questions. The analysis and discussion offer at least three open questions to be discussed and addressed in future studies: a) How can a computational approach capture all relevant semantic relations, going beyond just words, for better exploration of frames, 2) How can the semantic relations in one text document be connected with or informed by that of other documents for a broader understanding of frames across multiple documents, c) Given the role of many framing devices, not only words, in constructing frames (see Table 1), how can we develop a computational model that captures salience deployed through other framing devices including sentences, omit, metaphors, size and placement of texts, culture, emotion, sources, catchphrase, exemplars, visual content, etc.

A crucial part of framing analysis is to capture "how" a text is presented. Entman (1993)'s definition talks about "perceived reality" that also aligns with people's cognitive thoughts. In texts, the "perceived reality" is usually dissected between what is discussed and how it is framed. Though the "what" part is generally apparent, the main issue is to analyze the "how." In NLP, it appears difficult to automatically distinguish between the "what" and the "how." So, the framing analysis task in NLP is more complicated than for human analysts.

Limitations. Selecting articles for this survey was a challenging task as the words "frame" and "framing" are used in studies of other disciplines (e.g., engineering). This prompted us to exploit multiple ways (e.g., Google Scholar and Scopus) to collect relevant articles as comprehensively as possible. Articles not matching the keyword searches might have been left out. So, the list might have some articles missing due to the search constraints. We excluded non-English articles.

Regarding analysis, we mainly focused on methodological design and quality in terms of capturing and examining frames and framing devices. We did not focus and report on the accuracy of the models' performance. For example, we emphasized the quality of the training dataset (e.g., MFC) to explore frames, instead of the models' accuracies. As this survey article is conducted from a qualitative perspective, our results are constrained by quantitative insights (e.g., the frequency or percentage of applying particular methods in prior studies).

References

Lene Aarøe. 2011. Investigating frame strength: The case of episodic and thematic frames. *Political communication*, 28(2):207–226.

Afra Feyza Akyürek, Lei Guo, Randa Elanwar, Prakash Ishwar, Margrit Betke, and Derry Tanti Wijaya. 2020. Multi-label and multilingual news framing analysis. In *Proceedings of the 58th annual meeting of the association for computational linguistics*.

Christian Baden and Paul D'Angelo. 2018. Reconstructing frames from intertextual news discourse. *Doing news framing analysis II: Empirical and theoretical perspectives*, pages 43–66.

Monika Bednarek and Georgia Carr. 2021. Computer-assisted digital text analysis for journalism and communications research: Introducing corpus linguistic techniques that do not require programming. *Media International Australia*, 181(1):131–151.

Vibhu Bhatia, Vidya Prasad Akavoor, Sejin Paik, Lei Guo, Mona Jalal, Alyssa Smith, David Assefa Tofu, Edward Edberg Halim, Yimeng Sun, Margrit Betke, et al. 2021. Openframing: Open-sourced tool for computational framing analysis of multilingual data. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 242–250.

David M Blei. 2012. Probabilistic topic models. *Communications of the ACM*, 55(4):77–84.

Porismita Borah. 2008. Examining media content: A case study of newspaper coverage of dowry in india, 1999–2006. *Asian Journal of Communication*, 18(4):379–395.

Amber E Boydstun, Dallas Card, Justin Gross, Paul Resnick, and Noah A Smith. 2014. Tracking the development of media frames within and across policy issues. *Technical report, University of California, Davis*.

Amber E Boydstun, Justin H Gross, Philip Resnik, and Noah A Smith. 2013. Identifying media frames and frame dynamics within and across policy issues. In *New Directions in Analyzing Text as Data Workshop, London*.

Bjorn Burscher, Rens Vliegenthart, and Claes H de Vreese. 2016. Frames beyond words: Applying cluster and sentiment analysis to news coverage of the nuclear power issue. *Social Science Computer Review*, 34(5):530–545.

Pere-Lluís Huguet Cabot, Verna Dankers, David Abadi, Agneta Fischer, and Ekaterina Shutova. 2020. The pragmatics behind politics: Modelling metaphor, framing and emotion in political discourse. *ACL Anthology*.

Dallas Card, Amber Boydstun, Justin H Gross, Philip Resnik, and Noah A Smith. 2015. The media frames corpus: Annotations of frames across issues. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 438–444.

Paul DiMaggio, Manish Nag, and David Blei. 2013. Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of us government arts funding. *Poetics*, 41(6):570–606.

Paul D'angelo. 2018. Doing news framing analysis ii. *Empirical and Theoretical Perspectives*.

Robert M Entman. 1993. Framing: Towards clarification of a fractured paradigm. *McQuail's reader in mass communication theory*, pages 390–397.

Robert M Entman. 2003. Cascading activation: Contesting the white house's frame after 9/11. *Political Communication,*, 20(4):415–432.

Robert M Entman, Jörg Matthes, and Lynn Pellicano. 2009. Nature, sources, and effects of news framing. In *The handbook of journalism studies*, pages 195–210. Routledge.

Gail Fairhurst and Robert Sarr. 1996. *The art of framing*. San Francisco: Jossey-Bass.

Gail T Fairhurst. 2005. Reframing the art of framing: Problems and prospects for leadership. *Leadership*, 1(2):165–185.

Anjalie Field, Doron Kliger, Shuly Wintner, Jennifer Pan, Dan Jurafsky, and Yulia Tsvetkov. 2018. Framing and agenda-setting in russian news: a computational analysis of intricate political strategies. *arXiv preprint arXiv:1808.09386*.

William A Gamson, David Croteau, William Hoynes, and Theodore Sasson. 1992. Media images and the social construction of reality. *Annual review of sociology*, 18(1):373–393.

William A Gamson and Andre Modigliani. 1989. Media discourse and public opinion on nuclear power: A constructionist approach. *American journal of sociology*, 95(1):1–37.

Salma Ghanem. 1997. Filling in the tapestry: The second level of agenda setting in me mccombs, dl shaw & dh weaver (eds.), communication and democracy (pp. 3-15).

Fabrizio Gilardi, Charles R Shipan, and Bruno Wüest. 2021. Policy diffusion: The issue-definition stage. *American Journal of Political Science*, 65(1):21–35.

Guy Golan. 2021. What is news framing? an informal conversation among framing scholars. https://www.youtube.com/watch?v=mArApGS-p1I&t=57s.

Lei Guo, Chao Su, Sejin Paik, Vibhu Bhatia, Vidya Prasad Akavoor, Ge Gao, Margrit Betke, and Derry Wijaya. 2022. Proposing an open-sourced tool for computational framing analysis of multilingual data. *Digital Journalism*, pages 1–22.

James K Hertog and Douglas M McLeod. 2001. A multiperspectival approach to framing analysis: A field guide. In *Framing public life*, pages 157–178. Routledge.

Shanto Iyengar. 1994. *Is anyone responsible?: How television frames political issues*. University of Chicago Press.

Yangfeng Ji and Noah Smith. 2017. Neural discourse structure for text categorization. *arXiv preprint arXiv:1702.01829*.

Elise Jing and Yong-Yeol Ahn. 2021. Characterizing partisan political narrative frameworks about covid-19 on twitter. *EPJ data science*, 10(1):53.

Kristen Johnson, Di Jin, and Dan Goldwasser. 2017. Leveraging behavioral and social information for weakly supervised collective classification of political discourse on twitter. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 741–752.

Yowei Kang and Kenneth CC Yang. 2022. Communicating racism and xenophobia in the era of donald trump: A computational framing analysis of the us-mexico cross-border wall discourses: Special issue on donald trump era and communicating race in america. *Howard Journal of Communications*, pages 1–20.

Shima Khanehzar, Andrew Turpin, and Gosia Mikolajczak. 2019. Modeling political framing across policy issues and contexts. In *Proceedings of the The 17th Annual Workshop of the Australasian Language Technology Association*, pages 61–66.

Sotiris B Kotsiantis, Ioannis Zaharakis, P Pintelas, et al. 2007. Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering*, 160(1):3–24.

Haewoon Kwak, Jisun An, and Yong-Yeol Ahn. 2020. A systematic media frame analysis of 1.5 million new york times articles from 2000 to 2017. In *12th ACM Conference on Web Science*, pages 305–314.

Haewoon Kwak, Jisun An, Elise Jing, and Yong-Yeol Ahn. 2021. Frameaxis: characterizing microframe bias and intensity with word embedding. *PeerJ Computer Science*, 7:e644.

Pengxiang Li, Hichang Cho, Yuren Qin, and Anfan Chen. 2021. # metoo as a connective movement: Examining the frames adopted in the anti-sexual harassment movement in china. *Social Science Computer Review*, 39(5):1030–1049.

Siyi Liu, Lei Guo, Kate Mays, Margrit Betke, and Derry Tanti Wijaya. 2019. Detecting frames in news headlines and its application to analyzing news framing trends surrounding us gun violence. In *Proceedings of the 23rd conference on computational natural language learning (CoNLL)*.

Maxwell McCombs, Juan Pablo Llamas, Esteban Lopez-Escobar, and Federico Rey. 1997. Candidate images in spanish elections: Second-level agenda-setting effects. *Journalism & Mass Communication Quarterly*, 74(4):703–717.

Julia Mendelsohn, Ceren Budak, and David Jurgens. 2021. Modeling framing in immigration discourse on social media. *arXiv preprint arXiv:2104.06443*.

Nona Naderi and Graeme Hirst. 2017. Classifying frames at the sentence level in news articles. *Policy*, 9:4–233.

Thomas E Nelson, Rosalee A Clawson, and Zoe M Oxley. 1997. Media framing of a civil liberties conflict and its effect on tolerance. *American Political Science Review*, 91(3):567–583.

Viet-An Nguyen. 2015. *Guided probabilistic topic models for agenda-setting and framing*. Ph.D. thesis, University of Maryland, College Park.

Viet-An Nguyen, Jordan Boyd-Graber, Philip Resnik, and Kristina Miler. 2015. Tea party in the house: A hierarchical ideal point topic model and its application to republican legislators in the 112th congress. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1438–1448.

Tom Nicholls and Pepper D Culpepper. 2021. Computational identification of media frames: Strengths, weaknesses, and opportunities. *Political Communication*, 38(1-2):159–181.

Yotam Ophir, Dror Walter, Daniel Arnon, Ayse Lokmanoglu, Michele Tizzoni, Joëlle Carota, LORENZO D'Antiga, and Emanuele Nicastro. 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.

Stephen D Reese, Oscar H Gandy Jr, and August E Grant. 2001. *Framing public life: Perspectives on media and our understanding of the social world*. Routledge.

Margaret E Roberts, Brandon M Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana Kushner Gadarian, Bethany Albertson, and David G Rand. 2014. Structural topic models for open-ended survey responses. *American journal of political science*, 58(4):1064–1082.

Malik Sallam. 2021. Covid-19 vaccine hesitancy worldwide: a concise systematic review of vaccine acceptance rates. *Vaccines*, 9(2):160.

Lisa Sanderink. 2020. Shattered frames in global energy governance: Exploring fragmented interpretations among renewable energy institutions. *Energy research & social science*, 61:101355.

Antonio Sanfilippo, Lyndsey Franklin, Stephen Tratz, Gary Danielson, Nicholas Mileson, Roderick Riensche, and Liam McGrath. 2008. Automating frame analysis. In *Social computing, behavioral modeling, and prediction*, pages 239–248. Springer.

Mihai D Sturdza et al. 2018. Automated framing analysis: A rule based system for news media text. *Journal of Media Research-Revista de Studii Media*, 11(32):94–110.

Geoffrey Supran and Naomi Oreskes. 2021. Rhetoric and frame analysis of exxonmobil's climate change communications. *One Earth*, 4(5):696–719.

J Swenson. 1990. News coverage of the abortion issue. framing changes in the 1980s. paper presented to the committee on the status of women. *Association for Education in Journalism and Mass Communication*.

James W Tankard Jr. 2001. The empirical approach to the study of media framing. In *Framing public life*, pages 111–121. Routledge.

David Tewksbury and Julius Matthew Riles. 2018. Framing in an interactive news environment. *Doing news framing analysis II. Empirical and theoretical perspectives*, pages 137–162.

Isidora Tourni, Lei Guo, Taufiq Husada Daryanto, Fabian Zhafransyah, Edward Edberg Halim, Mona Jalal, Boqi Chen, Sha Lai, Hengchang Hu, Margrit Betke, et al. 2021. Detecting frames in news headlines and lead images in us gun violence coverage. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 4037–4050.

Wouter van Atteveldt and Tai-Quan Peng. 2018. When communication meets computation: Opportunities, challenges, and pitfalls in computational communication science. *Communication Methods and Measures*, 12(2-3):81–92.

TA Van Dijk. 2016. Analyzing frame analysis: A critical review of framing studies in social movement research. Technical report, Working paper version 4.0, 2 December. https://www. academia. edu/40286423 ....

Dror Walter and Yotam Ophir. 2019. News frame analysis: An inductive mixed-method computational approach. *Communication Methods and Measures*, 13(4):248–266.

Dror Walter and Yotam Ophir. 2021. Strategy framing in news coverage and electoral success: An analysis of topic model networks approach. *Political Communication*, 38(6):707–730.

Kenneth CC Yang and Yowei Kang. 2020. Framing national security concerns in mobile telecommunication infrastructure debates: A text mining study of huawei. In *Huawei goes global*, pages 319–339. Springer.

Tuukka Ylä-Anttila, Veikko Eranti, and Anna Kukkonen. 2021. Topic modeling for frame analysis: A study of media debates on climate change in india and usa. *Global Media and Communication*, page 17427665211023984.

Qi Yu. 2022. "again, dozens of refugees drowned": A computational study of political framing evoked by presuppositions. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Student Research Workshop*, pages 31–43.

Caleb Ziems and Diyi Yang. 2021. To protect and to serve? analyzing entity-centric framing of police violence. *arXiv preprint arXiv:2109.05325*.

A Appendix

Devices	Sources	Devices	Sources
Content (Texts)		**Content (Visual)**	
1. Words	Entman (1993)	25. Metaphors	Fairhurst and Sarr (1996); Gamson and Modigliani (1989); Tankard Jr (2001)
2. Stock phrases	Entman (1993)	26. Visual images (e.g., picture, icon)	Tankard Jr (2001); Gamson and Modigliani (1989)
3. Stereotyped image	Entman (1993)	27. Chart & graph	Tankard Jr (2001)
4. Sources of info	Entman (1993)		
5. Sentences	Entman (1993)	**Action**	
6. Metaphors	Fairhurst and Sarr (1996); Gamson and Modigliani (1989)	28. Placement (e.g., front page)	Entman (1993); Swenson (1990)
7. Jargon/catchphrase	Fairhurst and Sarr (1996); Gamson and Modigliani (1989),	29. Repetition	Entman (1993)
8. Contrast	Fairhurst and Sarr (1996)	30. Associating with culturally familiar symbols	Entman (1993)
9. Spin	Fairhurst and Sarr (1996)	31. Include	Entman (1993)
10. Stories	Fairhurst and Sarr (1996)	32. Omit or hide	Entman (1993)
11. Headlines & subheadlines).	Tankard Jr (2001)	33. Show root causes	Gamson and Modigliani (1989)
12. Subheads	Tankard Jr (2001)	34. Show effects	Gamson and Modigliani (1989)
13. Photo captions	Tankard Jr (2001)	35. Make appeals to principles (moral claims)	Gamson and Modigliani (1989)
14. Leads	Tankard Jr (2001)		
15. Selection of sources	Tankard Jr (2001)	**Context**	
16. Selection of quote	Tankard Jr (2001)	36. Contextual information	Baden and D'Angelo (2018)
17. Blown up quotes	Tankard Jr (2001)	37. Culture	Entman (1993)
18. Series' logos	Tankard Jr (2001)		
19. Statistics	Tankard Jr (2001)	**Communicator**	
20. Concluding statements	Tankard Jr (2001)	38. Thought	Fairhurst and Sarr (1996)
21. Exemplars	Gamson and Modigliani (1989)	39. Forethought	Fairhurst and Sarr (1996)
22. Depictions	Gamson and Modigliani (1989)	40. Being bias	Fairhurst and Sarr (1996)
23. Emotion	Aarøe (2011)		
24. Hashtag	Borah (2008)		

Table 1: Framing Devices Used to Construct Frame(s)

Citation	Type	Domain	Method/ Annotated corpora used	Resource
1. Boydstun et al. (2013)	Corpus, method	Tobacco, immigrant, same sex-marriage	Regression, Policy frames codebook (PFC)	N/A
2. DiMaggio et al. (2013)	Application	Artists & arts	Topic Modeling	N/A
3. Boydstun et al. (2014)	[1]	[1]	[1]	N/A
4. Card et al. (2015)	Method	[1]	Media frames corpus (MFC)	GitHub
5. Nguyen (2015)	Method	Congressional debates, reviews	Hierarchical topic modeling	GitHub
6. Nguyen et al. (2015)	Method	Congress speech	[5]	[5]
7. Burscher et al. (2016)	Application	Nuclear power	Cluster analysis	N/A
8. Ji and Smith (2017)	Application	Immigration	Neural network, semantic Relations	GitHub
9. Johnson et al. (2017)	Application	Abortion, affordable care act	[8]	GitHub
10. Naderi and Hirst (2017)	Application	Immigration, smoking	[8]	N/A
11. Field et al. (2018)	Application	U.S. coverage in Russian newspaper	[4]	N/A
12. Sturdza et al. (2018)	Method	N/A	Operationalization of semantic relations	N/A
13. Khanehzar et al. (2019)	Application	Immigration, same-sex marriage	[8]	N/A
14. Liu et al. (2019)	Method, annotated corpus	Gun violence	Gun violence frame corpus (GVFC), Neural network	GitHub
15. Walter and Ophir (2019)	Method	Senate coverage, epidemics	Topic modeling, Network analysis	GitHub
16. Akyürek et al. (2020)	Application & extension	[14]	[14]	GitHub1, GitHub2
17. Cabot et al. (2020)	Application	Immigration, smoking	[8]	GitHub
18. Kwak et al. (2020)	Application	Fake news	[4]	GitHub
19. Sanderink (2020)	Application	Renewable energy	Frequency and co-occurance model	Programs
20. Yang and Kang (2020)	Application	Telecom	[19]	N/A
21. Bednarek and Carr (2021)	Method	Lifestyle	[19]	WordsSmith
22 Bhatia et al. (2021)	Application & extension of open-source tool	Gun violence	[14]	[14]
23. Gilardi et al. (2021)	Application	Govt policy	Structured topic modeling	Appendix
24. Jing and Ahn (2021)	Application	Political polarization	FrameAxis	N/A
25. Kwak et al. (2021)	Method	Reviews	[24]	N/A
26. Li et al. (2021)	Application	#MeToo movement	[2]	N/A
27. Mendelsohn et al. (2021)	Application	Immigration	[8]	GitHub
28. Nicholls and Culpepper (2021)	Comparative	Banking	N/A	N/A
29. Ophir et al. (2021)	Application	COVID-19	[15]	N/A
30. Supran and Oreskes (2021)	Application	gun violence, oil and gas	[2]	N/A
31. Tourni et al. (2021)	Application & extension	Gun violence	[14]	[14]
32. Walter and Ophir (2021)	Application	[15]	[15]	[15]
33. Ylä-Anttila et al. (2021)	Application	Climate change	[2]	N/A
34. Ziems and Yang (2021)	Method	Police violence	Semantic relations	GitHub
35. Guo et al. (2022)	[22]	Gun violence	[14]	[14]
36. Yu (2022)	Method	Refugee crisis	[34]	GitHub
37. Kang and Yang (2022)	Application	Racism, Xenophobia	[19]	[19]

Table 2: Summary of the Methods and Resources