

News Kaleidoscope: Visual Investigation of Coverage Diversity in News Event Reporting

Aditi Mishra*

Shashank Ginjpalli†

Chris Bryan‡

Arizona State University

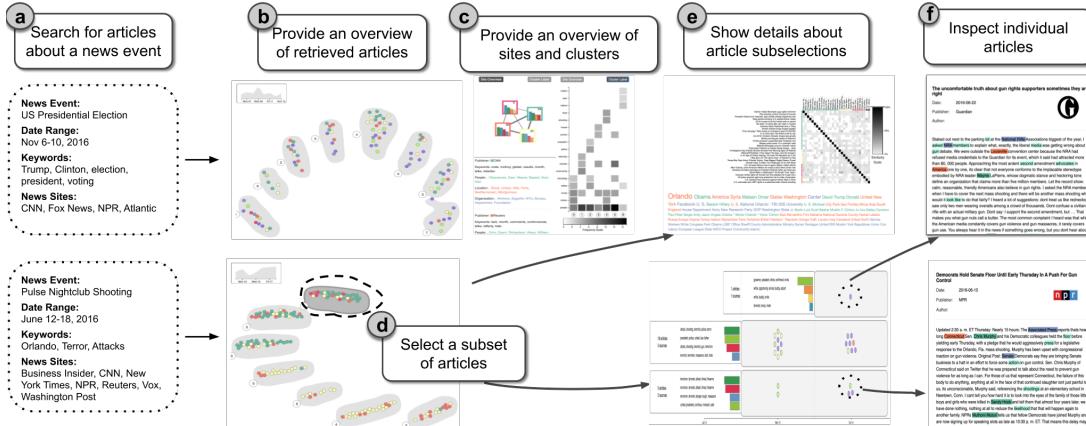


Figure 1: News Kaleidoscope supports visual analysis of discrete news events with a focus on how coverage of an event varies. (a) After searching for news articles about an event of interest, (b) an overview ordination plot shows the retrieved articles. (c) A site level view and annotations for rendered clusters is also displayed. (d) Selecting a subset of articles (e) populates subselection views, enabling further investigation of coverage diversity. (f) Individual articles can also be inspected for detailed review.

ABSTRACT

When a newsworthy event occurs, media articles that report on the event can vary widely—a concept known as *coverage diversity*. To help investigate coverage diversity in event reporting, we develop a visual analytics system called News Kaleidoscope. News Kaleidoscope combines several backend language processing techniques with a coordinated visualization interface. Notably, News Kaleidoscope is tailored for visualization non-experts, and adopts an analytic workflow based around subselection analysis, whereby second-level features of articles are extracted to provide a more detailed and nuanced analysis of coverage diversity. To robustly evaluate News Kaleidoscope, we conduct a trio of user studies. (1) A study with news experts assesses the insights promoted for our targeted journalism-savvy users. (2) A follow-up study with news novices assesses the overall system and the specific insights promoted for journalism-agnostic users. (3) Based on identified system limitations in these two studies, we refine News Kaleidoscope’s design and conduct a third study to validate these improvements. Results indicate that, for both news novice and experts, News Kaleidoscope supports an effective, task-driven workflow for analyzing the diversity of news coverage about events, though journalism expertise has a significant influence on the user’s insights and takeaways. Our insights developing and evaluating News Kaleidoscope can aid future tools that combine visualization with natural language processing to analyze coverage diversity in news event reporting.

Index Terms: Human-centered computing—Visualization—Visualization techniques—Treemaps; Human-centered computing—Visualization—Visualization design and evaluation methods

1 INTRODUCTION

In today’s digital society, the ways that people access news is rapidly changing. Consumption on traditional platforms (print, radio, and television) is falling as consumption via websites and apps increases [1]. In addition to legacy news organizations adopting online presences, new digital-native “born on the web” publishers are utilizing social media platforms such as Facebook, Twitter, and YouTube for outreach, engagement, and sharing of stories [1].

A corollary to this increasing media pluralism is that, when a newsworthy event occurs, articles reporting on the event will vary in different ways—a concept we refer to as *coverage diversity* [25]. While coverage diversity can include structural aspects—e.g. the language a news article is written in and its length—it also includes thematic and framing aspects such as media bias. Systematic analyses have identified multiple types of media bias. For example, AllSides, a media watchdog group, classifies media bias into eleven categories, including the use of spin words, flawed logic, sensationalism/emotionalism, and ad hominem attacks [4].

Our research aim is, given a news event of interest, to support the visual analysis the coverage diversity of news articles that report on it. Specifically, we consider coverage diversity in terms of the *keywords* and *entities* employed in reporting and their potential emotional biases, which are classified as emotionalism/sensationalism types of media bias [4]. We employ a design study methodology [43] to develop and evaluate a novel visual analytics platform, called News Kaleidoscope, which combines interactive visualizations and natural language processing (NLP) techniques to analyze coverage diversity.

News Kaleidoscope is designed based on a pre-study with jour-

*e-mail: amishr45@asu.edu

†e-mail:sginjpali@asu.edu

‡e-mail:cbryan16@asu.edu

nalism researchers. These users are interested in analyzing news reporting (and coverage diversity), but they are not visualization or NLP experts. To identify “polarities” in coverage diversity, we adopt a workflow around selecting article subsets (*subselections*), based on the premise that while articles about an event of interest will largely share top-level keywords and factoids, subselections will vary in second-level keywords, entities, and emotional stylings that enable a more nuanced understanding of coverage diversity.

To understand how News Kaleidoscope supports the analysis of coverage diversity, we conduct an extensive set of evaluations. We first evaluate News Kaleidoscope with the journalism-savvy target users (Study #1), whom we call *news experts*, to understand the specific types of insights promoted by the system. Then, to understand if and how News Kaleidoscope generalizes to a broader user audience, we conduct a follow-up study with *news novices* (Study #2). The participants in this study have little familiarity with U.S. news sites and therefore few preconceived assumptions about how coverage will differ when reporting on news events. Based on feedback from these studies, we implement a targeted set of design improvements to News Kaleidoscope and conduct a validation study (Study #3) to verify their efficacy.

Our results indicate that News Kaleidoscope effectively supports analysis of coverage diversity both for news experts and news novices, though interestingly it provides different benefits to each user group (e.g., for news experts, it supports validating hypotheses and assumptions about news sites). Based on our experiences in creating and evaluating News Kaleidoscope, we discuss generalizable takeaways for visually analyzing coverage diversity in a news landscape that is increasingly online and diversified. In summary, the contributions paper include: (1) we identify requirements for visually analyzing coverage diversity, which constitutes a current and real-world problem in journalism and media research, (2) we design and implement a novel visualization system, News Kaleidoscope, for analyzing news articles about an event of interest in the context of coverage diversity, (3) based on a series of robust evaluations with both news experts and novices, we learn about design guidelines and implications for analyzing coverage diversity using visualization and NLP.

2 RELATED WORK

Bias and Framing in News Reporting. Biased reporting can have substantial societal impacts. For example, a 2007 study by DellaVigna and Kaplan [14] found evidence that media bias had a significant effect on voting in the 1996 and 2000 presidential elections. More recently, a 2015 study [22] showed that exposure to biased news can lead to intolerance of dissent and group polarization.

Today, watchdog organizations such as AllSides [2] and the Center for Media and Public Affairs [3] assess media bias and framing, primarily via qualitative assessments. In contrast, academic research has employed data-driven approaches to quantify bias in news reporting. A seminal 2005 paper by Groseclose and Milyo [23] assigned bias scores to news organizations by counting think-tank citations in articles, scoring the majority of studied outlets as left-leaning. A recent survey paper [24] reviews several automated approaches for identifying media bias in news articles, including ones that blend NLP with visualization, and recent work has investigated the use of machine learning for identifying and classifying bias (e.g., by analyzing the propaganda techniques used by news agencies [13]).

Visualizing News Reporting and Text Data. News visualization tools support the analysis of reporting. For example, LingoScope [16] and Compare Clouds [15] enable the comparative analysis by visualizing keywords distributions between media groups or news sites. TimeMines [46] generates event timelines based on semantic features such as keywords and entities present in the text of articles (similarly, Chieu and Lee [10] support time-based event extraction via keyword queries). While News Kaleidoscope is a

news visualization tool, its intended goal differs from these systems.

News-focused interfaces build upon the broader community of visualization of text data. Such visualizations normally leverage text mining and/or NLP techniques, including keyword extraction, entity recognition, event and topic modeling, sentiment analysis, and document similarity/clustering [8, 28, 31], to transform unstructured text corpora into derived text (meta)data that is suitable for visual analysis. Several general-purpose visualization systems have been designed for high-level summarization and browsing of events/topics via aggregate visualizations [12, 18, 33]. Alternatively, a collection of documents can be plotted as discrete data points using techniques such as matrices [5], clustering [45], parallel coordinates [17], force-directed layouts [29], and dimensionality reduction or ordination [9, 26]. Several of these techniques require computing the pairwise similarities between documents—see Cao and Cui for an overview [8]. In News Kaleidoscope we employ a multi-weight aggregate distance metric that can be interactively adjusted.

Contextualizing News Kaleidoscope to Previous Work. In contrast to prior-mentioned news and text visualization systems, which support tasks such as the broad summarization of events over time [33] or support comparing the keywords used by two media sites [16], News Kaleidoscope’s design is tailored towards a specific demographic and goal: people who are interested in analyzing the coverage diversity of news reporting, but who are likely not visualization or NLP experts. In contrast to general-purpose visualization interfaces for text and document corpus analysis, news articles have special semantics—for example, articles reporting on a news event likely share similar top-level keywords and emotions, but it may be necessary to investigate second-level nuances in the data to understand if and how coverage diversity is occurring. Further, we carefully design News Kaleidoscope to balance its analytical capabilities with a user experience and visualization designs that are accessible to the target user base, by following a rigorous design study methodology [43] that results in generalizable takeaways demonstrating how the visual analysis of coverage diversity differs from general text visualization, particularly when accounting for the user expertise.

3 DESIGN REQUIREMENTS ANALYSIS

To motivate a visual analytics design for investigating coverage diversity, we interviewed a trio of news experts: journalism professors who research or teach on news reporting and media (two experts are currently assistant professors, and the third is a full professor).

A significant problem in the journalism community (stated by each participant) is that there are a lack of computational tools or visualization software specifically designed for analyzing of reporting various styles or themes (i.e., coverage diversity) in news-based reporting. While the participants knew computational processes existed for analyzing text data (e.g., NLP algorithms), they had little familiarity with the technical aspects (leading to high uncertainty and limited trust). Based on these interviews, we derived a set of five design requirements (**DRs**) to support visual analysis specifically in the context of analyzing news reporting for coverage diversity. For each DR, we provide a brief justification about how it supports the intended domain (i.e., visualization non-experts who want to analyze news articles). Section 4 describes the software stack in detail.

DR1: Retrieve articles about an event of interest. The common “first step” for our participants in their analytic workflows was to search for news articles about a topic of interest via keyword-based queries in search engines, then skimming or reading through the text of retrieved articles. This workflow lies in contrast to some previous systems for document visualization, which begin by showing an aggregate view of the entire collection (e.g., [18, 33]). Since such a summarization perspective is extraneous for the current domain (i.e., the participants already know what topic they want to search for, and what its high-level keywords are), we omit a summary view

and instead allow users to directly query for a topic of interest based on high-level keywords and other constraints (date, source, etc.).

DR2: Provide a high-level overview of coverage diversity. Participants normally reviewed the retrieved articles about a news topic in an iterative manner (e.g., as a list), which could quickly become overwhelming. The participants described that *providing an overview of the retrieved articles about the news event of interest, in a way that emphasises the general (i.e., high-level) trends or groupings of the coverage diversity, would provide an initial sense of the event’s coverage diversity*. Such a view would also function to drive subsequent analysis, via an overview-plus-details workflow [44].

DR3: Select subsets of articles to analyze coverage diversity polarities. In addition to providing high-level overview of an event’s coverage diversity, participants described wanting to investigate specific polarities in reporting—subsets of articles containing interesting biases or semantics compared to the rest of the reporting articles. Analyzing the second-level keywords in these polarities, as well as the presence of polarity-specific emotional media biases, would provide nuanced understanding of the coverage diversity for a news event. Participants also desired that such analyses be able to compare both within and across article subsets. In other words, *selecting article subsets should enable detailed analysis of coverage diversity polarities via tailored visualizations*.

DR4: Provide data-level explanations to ensure trust and verification. For users not familiar with complex NLP and/or data mining processes, it is important to provide trust and interpretability of models and algorithms to non-expert users [30]. In our discussions about these topics, our participants were wary about simply trusting the outputs of algorithms or models they had little insight into. For example, in discussing the clustering of news articles, participants wanted to know why an articles might be binned a certain way. Thus, *when advanced computational techniques or models are employed, mechanisms should be employed to promote trust and interpretability of model recommendations and decisions*.

DR5: Visualization complexity should account for user expertise. Like the majority of journalism researchers, our participants were not experts in advanced visualization interfaces. This means that overly complex or esoteric visual designs could lead to a failure in conveying information succinctly and easily. Instead, *interface designs should strike a balance between providing sufficient analytic capabilities while also being approachable and intuitive to use*. News Kaleidoscope is built with these considerations in mind, and validated in the user studies described in Section 6.

4 THE NEWS KALEIDOSCOPE SYSTEM DESIGN

We now describe News Kaleidoscope’s system design. The system is a full-stack application with three primary facets: (1) a data preprocessing step, (2) a backend server for data storage, query, and NLP-based computation, and (3) a frontend interface for visualization and interaction. For examples of how News Kaleidoscope can be used to analyze coverage diversity, see the use case in Section 5 and the demo video include in the supplemental materials.

4.1 Data Corpus and Preprocessing

When describing News Kaleidoscope, we use a large news article dataset titled *All the News* [47]. This text corpus contains 143,000 articles published on 14 news sites from 2015–2017 and includes both the article text and related metadata (title, news site name, author, publication date, article URL, etc.). The news sites are Western media organizations (the only non-American site is The Guardian) across a spectrum of perceived liberal-to-conservative political biases. Figure 4 includes a list of the 14 news sites and labels their “media bias ratings” as scored by the AllSides organization [2]. By using this diverse corpus as a dataset, we ensure that reporting on news events will likely have a high degree of coverage diversity.

For each article in the corpus, we apply a trio of heuristics as pre-processing steps: keyword identification, named entity recognition, and determination of emotional style. **Keywords** from an article are extracted via the Gensim NLTK library [32]. **Named entity recognition** follows a similar process. We use the Stanford NER library [20] to extract named persons, locations, and organizations in each article. These are stored to a set of three vectors for each article: one for names, one for locations, one for organizations. Vectors are extracted using the bag of words technique for each article retrieved, where the presence or an absence of a word is encoded as a 1 or a 0, respectively. While keywords and entities might at first glance seem redundant, there are important semantic differences. Keywords give a broad sense of the topical content of an article, but the individual words are unclassified. In contrast, entities are specifically binned into classes, providing the user with an explicit set of relevant, descriptive proper nouns.

The third heuristic characterizes the **emotional style** of each article. We use Plutchik’s discrete categorical model [39] to classify emotional states into eight primary emotions, organized as the following pairs: anger-fear, anticipation-surprise, joy-sadness, and trust-disgust. The NRC lexicon [38] (which is based on Plutchik’s model) extracts and classifies emotional words contained in an article. We then compute the frequency of words for each emotion in the article. Specifically, we construct a 1×8 vector (e_1, e_2, \dots, e_8) . Each e_i equals n_i/N , where n_i is the number of words for that particular emotion, and N is the total number of words in the article (after stop-word removal). This means each article has an emotional vector of length = 8, where each e_i represents how much of that particular emotion exists in the article’s text. Using this multi-dimensional characterization enables a much more robust analysis compared to simpler positive/negative sentiment models (e.g., the popular Stanford CoreNLP library [36]).

4.2 Backend Server

The backend server is built using Node.js [48] and acts as the storage and service layer between the processed data and the frontend interface. News articles—along with extracted keywords, entities, and emotional style vectors—are stored to a SQL database.

Based on a user query—which can include constraints such as date, keywords, news sites to include/exclude, and the number of articles to retrieve—we retrieve articles from the database (**DR1**) and rank them using TF-IDF [41]. Specifically, we consider the query as a single small document and calculate the TF-IDF for each article in the database that meets the date/site constraints, using this to sort the articles by relevance. The desired number of articles are then returned. For this returned collection of articles, we compute their aggregate pairwise distances for the overview visualization in the interface (**DR2**). This heuristic is a combination of three independent distance metrics: the preprocessed (1) keyword and (2) entity vectors, and also (3) the temporal similarity of articles based on their publication dates.

Keyword Distance. For all articles returned in a query, we compute the pairwise Jaccard similarity between their keyword vectors, normalized over the total number of words in each pair of articles. For articles a_1 and a_2 , their keyword distance $d_k(a_1, a_2)$ represents how similar they are in terms of keywords on a scale between [0, 1].

Entity Distance. Similar to keyword distance, for each article pair we compute the normalized Jaccard similarity combined from each of three entity vectors (name, location, and organization), and then normalize the value over the total entities in the paired documents. Thus, the entity distance between articles a_1 and a_2 is given by $d_e(a_1, a_2)$, again on a scale from [0, 1].

Temporal Distance. Given that a news event generally happens over a discrete timeframe, we assume that reporting articles published in close temporal proximity (such as on the same day) might be more similar, as news coverage tends to evolve as follow-up

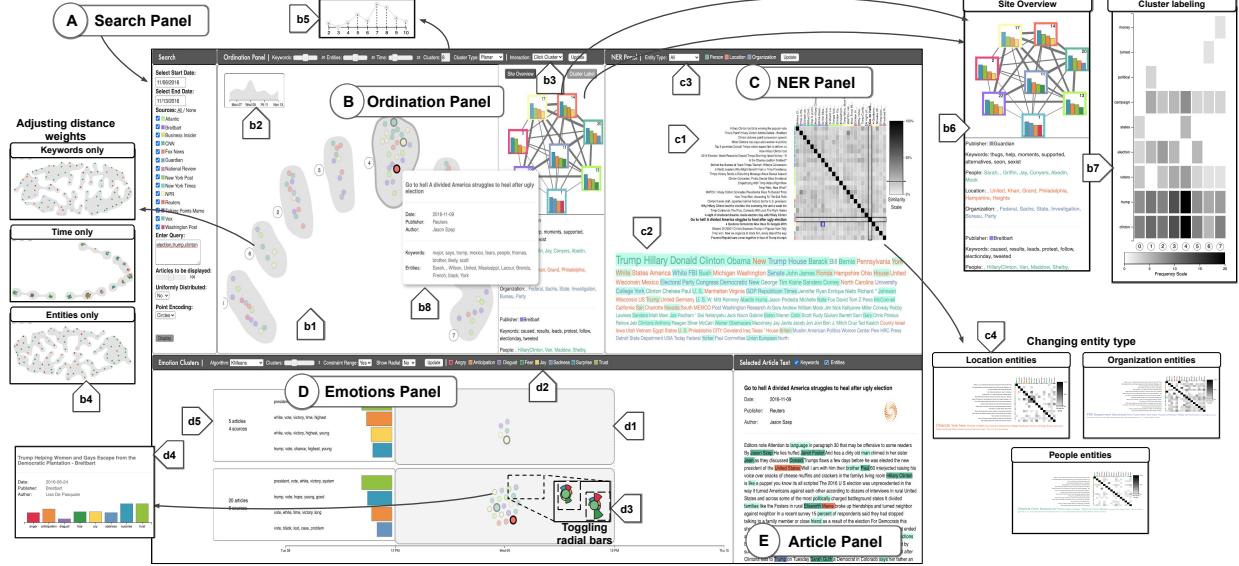


Figure 2: The News Kaleidoscope interface consists of five panels to support (A) article search, (B) overview visualizations, (C–D) detailed analysis, and (E) inspection of individual articles.

investigation and reporting occurs [50]. Given a time range R from which a set of articles are queried, we compute the temporal distance between articles a_1 and a_2 as $1 - \frac{\text{date}(a_1) - \text{date}(a_2)}{R}$. This distance $d_t(a_1, a_2)$ has a range between $[0, 1]$.

Aggregate Pairwise Distance. For a set of retrieved articles, the pairwise keyword, entity, and temporal distances are independently computed and combined into a single aggregate distance $dist(a_1, a_2)$ which represents the overall similarity between two articles:

$$dist(a_1, a_2) = w_k \times d_k(a_1, a_2) + w_e \times d_e(a_1, a_2) + w_t \times d_t(a_1, a_2)$$

Each distance metric is multiplied by a scaling weight (w_k , w_e , and w_t); these are interactively adjustable in News Kaleidoscope’s frontend interface, enabling a multi-faceted exploration of articles based on desired user semantics. For example, setting $w_k = 0$ would mean that the keyword distance would have no effect on the aggregate distance. One advantage of this aggregate multi-weight heuristic is that new similarity metrics can easily be added as desired. For example, a new metric for “author similarity” could be computed by analyzing the historical style or content of a writer.

Clustering Articles by Emotional Style. When article subselections are made (**DR3**), News Kaleidoscope clusters these articles by their emotional styles. To do this, we calculate the pairwise similarity in emotional style for articles, and then cluster the articles using k -means clustering [35]. The similarity is the Euclidean distance using the precomputed 1×8 emotional vector.

4.3 Frontend Interface

Figure 2 shows an overview of News Kaleidoscope’s frontend interface. It consists of five linked panels (**A–E**), designed to support the tasks #1–6 described in Section 3. As a note, this image shows the final design of News Kaleidoscope that includes four additional features added based on participant feedback from Studies #1 and #2—(b5), (b6), (b7), and (d5)—which were evaluated in Study #3.

(A) Search Panel. The search panel supports keyword-based searches for articles about news events (**DR1**). The user can set constraints/filters for date ranges, news sites, keywords, and the number of articles to retrieve. The user also has the option to return a uniform number of articles for each new sites: e.g., if 140 articles are retrieved for 14 sites, each site will return 10 articles. Such a

constraint provides a “balanced” distribution of article sources, but has the potential to lead to “less relevant” results if a news event was not extensively covered by a site.

(B) Ordination Panel. The ordination panel provides an overview of retrieved articles (**DR2**). (b1) Articles are encoded as circles, and colored by news site, and laid out via dimensionality reduction using aggregate pairwise distances. News Kaleidoscope supports layout via multidimensional scaling (MDS) [27], t-SNE [34], and UMAP [37]. For the user studies, we employ MDS as it results in deterministic layouts, thus eliminating a potential confounding variable. Circles are clustered via k -means clustering, with cluster hulls rendered using bubble sets [11]. We are motivated to employ this approach as it provides a scalable view of the retrieved articles, intuitively using proximity and groupings between data points to indicate their similarity. (b2) At upper left, an area chart shows the temporal distribution of retrieved articles.

(b3) At top, several widgets control display settings, including selecting the distance metric for cluster computation (either aggregate pairwise distance or x/y positioning), and updating the k value to change the number of clusters. A trio of sliders adjust the scaling weights for the aggregate pairwise distances (the w_k , w_e , and w_t values), which updates the layout in MDS plot (b4). (b5) Hovering on the k value input shows a tooltip that shows the silhouette score [42] for each value of k between 2 and 10 (a higher score indicates better separation between clusters), as a way to explain to non-technical users what constitutes “good” choices for k .

Two additional overview plots are shown in (b6) and (b7). These two displays (placed in a tabbed panel) support comparative analysis at the news site-level and between clusters. (b6) The site overview tab shows a graph of all news sites with retrieved articles; edges indicate the overall keyword and entity similarity between the content of their stories. To provide a glimpse of each site’s reporting, the top four emotions for each site are plotted in a bar chart (labeled by total stories for each site, with bar colors based on the emotions panel’s color palette (d2)). This view provides a summary-level comparison of how news sites differ in their reporting on the current event. The bottom half of this tab shows the most popular keywords and entities for each site.

(b7) The cluster labeling contains a heatmap showing (in rows) the top keywords from each cluster, while (in columns) clusters

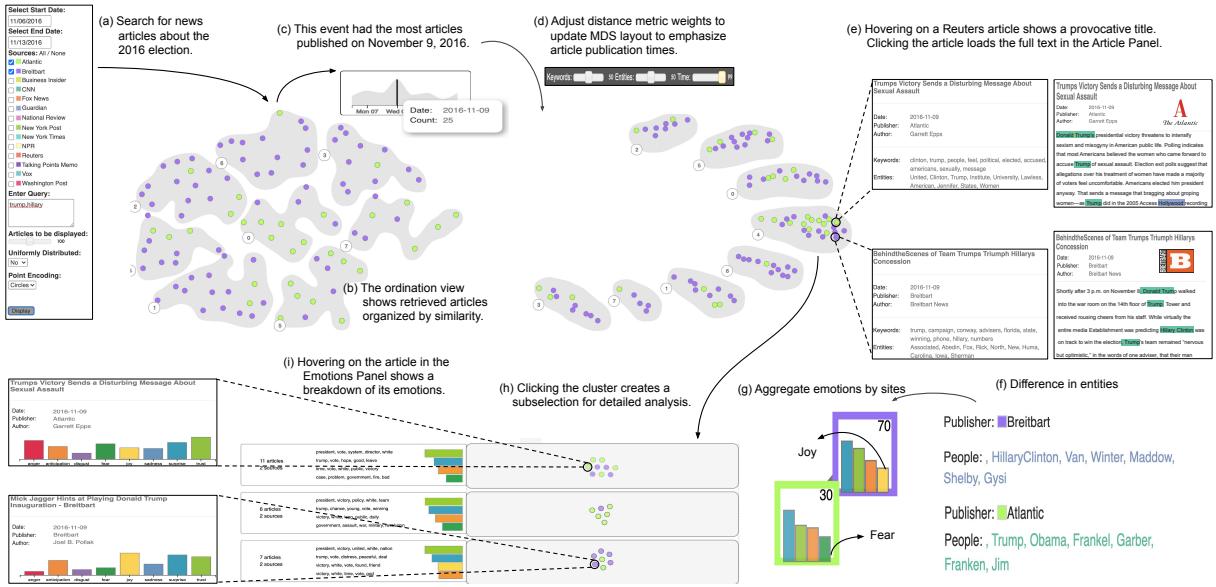


Figure 3: The interactions taken during the Use Case Scenario described in Section 5.

are columns ordered by index. Cell color indicates the frequency that top keywords appear in each cluster, supporting analysis of high-level coverage trends across the clusters.

(b8) Finally, hovering over an article displays its tooltip. To select a subset of articles from the overview (**DR3**), there are two available interactions. Drawing a lasso makes a freeform selection, while clicking on a cluster selects all of its articles. When an article subselection is made, the NER and emotion panels are populated for detailed analysis.

(C) NER Panel. The NER panel visualizes the article subselection with (c1) an adjacency matrix and (c2) word cloud. The adjacency matrix shows pairwise article similarities between articles based on the entity distance metric, allowing for direct comparison of two articles. (c3) The user can choose which named entity types are used to compute similarity: persons, locations, organizations, or all (c4). Entity distance is used for this visualization (instead of keywords) as it bins extracted article text into explicit categories. Below, a word cloud shows which entities are most used in the article subselection. Words are ordered by frequency and colored by entity type. Hovering on a cell in the adjacency matrix highlights shared entities in the word cloud.

(D) Emotions Panel. The emotions panel visualizes the emotional styles of the article subselection. Articles are clustered via k -means clustering using the emotional style vectors. (d1) Articles within each cluster are temporally ordered according to their publication day. (d2) Using the panel's control widgets, cluster settings can be updated and article circles can be toggled to display as radial bar charts (d3). This view shows each article's emotional style vector as a set of eight radial bars around its circle. (d4) Hovering on an article shows this vector using a (standard) bar chart. (d5) At the left of each cluster, we show that cluster's top-4 dominant emotions (as well as the top contributing words to each emotion) to provide a glimpse of the cluster's overall characterization.

(E) Article Panel. The article panel allows inspection of individual articles (**DR5**) by showing the raw article text and metadata (author, news site, publication date). Articles are loaded in this panel by clicking on an article circle (ordination and emotion panels) or adjacency matrix title (NER panel). To support analysis, the user can highlight the extracted keywords and entities in the article's text.

5 USE CASE SCENARIO

To illustrate how News Kaleidoscope can be used to analyze coverage diversity, we present a use case scenario with Gary, a journalist who is reviewing reporting on the 2016 U.S. presidential election. This event, which took place on November 8, 2016 between Democrat Hillary Clinton and Republican Donald Trump, had reporting both leading up to the day as well as post hoc analysis and reporting on the outcome. Figure 3 shows his workflow.

(a) Gary first searches for articles from two media organizations (Breitbart and The Atlantic) using the keywords *trump* and *hillary* from November 6–13, 2016. (b) The retrieved articles populate in the ordination panel. (c) By reviewing the temporal distribution of articles, Gary notices most are published on November 9 (the day after the election). (d) Gary increases the weight of the temporal distance metric, which updates the MDS layout to place articles published on the same day in the same cluster. (e) Using the tooltip to skim articles in the November 9 cluster, Gary notices that green circles (Atlantic articles) emphasize the *trump* keyword, however they seem to be downplaying his electoral win (example articles include *Trump's Victory Sends a Disturbing Message About Sexual Assault* and *Empathizing with Trump Voters Right Now*). In contrast, Breitbart articles (in purple) focus on Hillary Clinton's loss (e.g., *Justice has Prevailed with Hillary loss* and *Behind-the-Scenes of Team Trump's Triumph, Hillary's Concession*). (f) This is further validated by reviewing the Site Overview Panel; the People entity for The Atlantic focuses on Trump, while Breitbart focuses on Clinton. This suggests that the two sites are exhibiting coverage diversity in how they are covering this event.

Gary now wants to understand *how* these two sites are differing in their coverage. (g) He uses the Site Overview Panel to review the aggregate emotions expressed by these two sites. Three of the top four emotions (surprise, trust, anticipation) are shared between the sites, however The Atlantic includes fear and Breitbart includes joy as an emotion. (h) To analyze the emotions in depth, Gary selects the November 9 cluster as an article subselection into the Emotions Panel. He notices an interesting grouping of articles across this panel: the first and second clusters consist primarily of Atlantic articles, and one of the dominant emotions in these clusters is fear (this is supported by the highlighting of keywords in the Article Panel such as *assault*, *war*, *military*, and *revolution*). The third cluster, which

has high Breitbart representation, shows joy as a dominant emotion (with keywords such as *victory*, *friend*, and *God*). (i) Hovering on individual articles further demonstrates this: Atlantic articles tend to have a higher amount of words associated with emotions like anger and fear.

Gary has now confirmed both that coverage diversity exists in how The Atlantic and Breitbart have covered this event, and he has learned how the polarity has manifested in terms of emotional biases and keywords. The Atlantic has tended towards coverage that evokes anger and fear in the context of Trump's win, while Breitbart shows joy regarding Hillary Clinton's loss.

6 EVALUATION

To empirically evaluate News Kaleidoscope, we conducted a trio of user studies. Study #1 consists of pair analytics sessions with 7 news experts (e1–e7), who are the target demographic for the system. However, to understand if and how News Kaleidoscope can generalize to a broader population, we subsequently conducted a follow-up study with 11 news novices (p1–p11). In contrast to Study #1, we format Study #2 as a controlled usability study to understand not only the insights supported by News Kaleidoscope, but also the overall user experience for users with little background domain knowledge.

These two studies provided robust insights into how News Kaleidoscope supports analysis of news events and coverage diversity, but they also highlighted several system features that were either missing or were difficult to understand. After adding a targeted set of new system features based on this feedback, we conducted a third study to validate their utility (Study #3).

6.1 Study #1 Design

For Study #1, we conducted pair analytics sessions with 7 news experts. Specifically, we wanted to understand what specific types of insights about coverage diversity were supported by News Kaleidoscope. Pair analytics [6, 19] is a procedure for capturing reasoning processes in visual analytics. A visualization expert well-versed with an interface's functionality "drives," while the study participant reasons and makes decisions based on their domain expertise. Verbal discussion between the driver and participant is the basis for understanding of the participant's sensemaking process as well as what specific insights are uncovered during investigation.

Study Procedure. Study sessions were conducted remotely using Zoom videoconferencing software. Kaleidoscope's browser window was shared via Zoom's sharing feature from the visualization driver's computer to the remote participant. QuickTime Player recorded audio discussion and screen capture.

For each session, the visualization expert first gave an overview of the system's functionality until the participant felt confident to proceed. After this, the news expert could freely analyze and explore the system using the entirety of the *All the News* text corpus. Sessions lasted as long as the news expert desired, though it was suggested they spend at least 15 minutes in their investigation. At the end of each session, participants were asked to provide freeform commentary and feedback on the system, such as what features they liked, disliked, and found useful.

Participants. The 7 participants in Study #1 were professors at Arizona State University in the political science and journalism departments (5 assistant and 2 associate professors). All participants were highly knowledgeable about American news reporting, including coverage diversity and perceived biases of news sites; a couple of the faculty specifically researched media bias and coverage diversity in reporting.

6.2 Study #1 Results

To understand how News Kaleidoscope promotes insights to news experts, we reviewed the discussions from recorded screencasts. We report here the usage patterns and the insight themes that were promoted.

News Kaleidoscope supports scalable analysis of news events. An immediate recognized benefit, noted by all participants, was that News Kaleidoscope quickly allows them to see and analyze a large number of stories about a news event: "*The fact that there are so many articles and site at the same place makes it useful already*" (e1). "*The system would be a good place to start for journalists because we deal with a lot of news articles and it would definitely be easier if such a system would be used*" (e2). "*Overall I would use this system to start with a big dataset analysis because it seems really good for that*" (e5). "*This system would definitely be a great place to start. So many articles and news sites in one place. I find this system really comprehensive*" (e3).

Multiple participants quickly realized that articles could be ingested at by reviewing keywords instead of reading the entire text: "*Visualizing keywords and entities like this is really great. I don't have to read the entire article at all*" (e4). "*The keywords and entities are a good thing in the system, it very nicely shows me which article is similar and which is it different from*" (e1).

Several participants mentioned that News Kaleidoscope showed coverage diversity through the lens of *polarity* between news sites and over time. "*Bias isn't explicitly shown, but as a user walks through or uses the system more I think it would be easier to see polarities using the emotion panel ... the system gives me a very good idea of coverage*" (e1). "*This would help me a lot in the kind of work I am doing. I work on the usage of words in a news article to then find the underlying polarity ... this system would be just apt for it*" (e4).

The visual analytics were a novel experience. While one participant (e6) had previous experience using LDA [7], no participants had used a visual analytics system similar to this as a part of their own work or research. "*I have never used such a system before, I have done topic modeling, but this is interesting and new*" (e6). "*I think just the fact that so many features and site comparison exists in the system makes it interesting because we generally don't have anything like this to use*" (e2). "*I really like the emotions panel. It's a very new take and it shows me a broad spectrum over time, which is really informative for me and my research*" (e5).

An unexpected usage of News Kaleidoscope as a validation system. While data exploration and discovery are common visualization goals, empirical *validation*—such as of a hypothesis or model's performance—is also a function that visualization can support. Since news experts had existing knowledge and intuitions of coverage diversity (as well as the potential biases of news sites), several participants felt the system could be used as a verification tool. "*Since I already know the bias or the kind of articles sites publish, I might use this system to compare them and validate my understanding*" (e2). Interestingly, e7 mentioned using the system as a way to review news articles that she helps create: "*We can use it to understand what's going on: model a [news] outlet. It might be helpful for comparison. How are we doing in presenting news? Are we similar to CNN or NPR?*"

Targeted suggestions by news experts. While participants could understand the visualizations and interactions, some took a bit of time to acclimate to the workflow and the system's available functionalities. "*Adding a few breadcrumbs might help, because they are so many things going on here*" (e5). Likewise, as non-engineers, participants were sometimes unsure of the backend algorithms. For example, clustering was positively reviewed, but several participants (e1, e2, e4, e5) suggested adding some form of explanation as to how the metrics were computed; e.g., why the specific number of clusters was chosen: "*A little extra support for people with no CS background would help, because its a very new thing for me*" (e4). "*A lot of non-computer scientist people might find it a little overwhelming*" (e3). Another feature that was requested multiple times was summary and comparative views, to show aggregate event behavior and site-level comparisons: "*I preferred more overview*

... [Start by showing] just showing one cluster first and then the user digs in deeper” (e5). “I wished it directly showed me the sites’ differences or similarities” (e3).

6.3 Study #2 Design

Study #1 provided understanding into how News Kaleidoscope can be used by domain experts to analyze coverage diversity. However, during development we also wondered if such a system can benefit news novices—users that are largely agnostic about the coverage diversity of specific news sites. As a motivation for this, a recent RAND report indicates that people are increasingly searching for “alternative views” about newsworthy events of interest [40]. Therefore, we conducted a second study targeting these news novice users.

In contrast to Study #1, and to better regulate the user experience and understand the thought process of news novices, Study #2 was run as a full qualitative study that included both in-study think-aloud protocol and a post-study questionnaire and feedback session about the user experience. This allows us to understand both how the insights of novice users compare and contrast to domain experts, and also to understand the overall system experience for a set of domain-agnostic users.

To begin Study #2, participants first completed a survey to collect demographic information and background knowledge. After this, the following three stages were run:

Training Stage. Each participant was given a hands-on tutorial describing all features and interactions in News Kaleidoscope. Participants could then practice with the system for as long as desired, analyzing news events in the *All the News* corpus that occurred in January 2017. A list of four noteworthy events was provided to the participant as a reference point. The intent was for the participant to become comfortable using the interface before proceeding to the main study.

Exploration Stage. Next, each participant was free to explore the news corpus using News Kaleidoscope for as long as they desired. It was suggested that participants spend at least 15 minutes using the system, however there was no hard cutoff. To constrain the study design, we restricted article queries to news events that happened in the latter half of 2016 (approximately 62,700 total articles). Participants were instructed to start by exploring and analyzing the coverage of two significant news events from that year: (1) The 2016 presidential election on November 8. (2) The Pulse Nightclub shooting in Orlando, Florida on June 12. After, participants were free to explore and analyze any news events they wanted. A list of noteworthy news events (with associated keywords) was provided for reference, though participants could query and use the system however they desired. During this stage, participants used think-aloud protocol to verbalize their thought processes and actions [21].

Review Stage. Finally, participants completed a short survey questionnaire about system impressions and functionality using a Likert scale (1 – strongly disagree, 7 – strongly agree). Participants then had the opportunity to provide freeform comments, suggestions, and criticisms about the interface and their experience.

Participant Recruitment and Apparatus. We recruited 11 non-domestic engineering students from <Anonymous University>; age $\mu = 24.27$ years ($\sigma = 1.95$), 7 males and 4 females, most of whom (8/11) had only moved to the United States after 2016. All participants were proficient in English and had good (corrected if necessary) eyesight. News Kaleidoscope was displayed in Google Chrome on a 24-inch monitor (3840×2160 resolution) with keyboard and mouse and connected to a MacBook Pro running macOS Mojave. QuickTime Player recorded session screencasts and audio.

To determine if Study #2 participants could be considered news novices, we included a pre-screening as a part of the background knowledge collection. As all participants were non-domestic students, this included asking several questions about American news reporting. While many participants reported some familiarity with American news companies based on a 7-point Likert scale ($\mu = 4.58$,

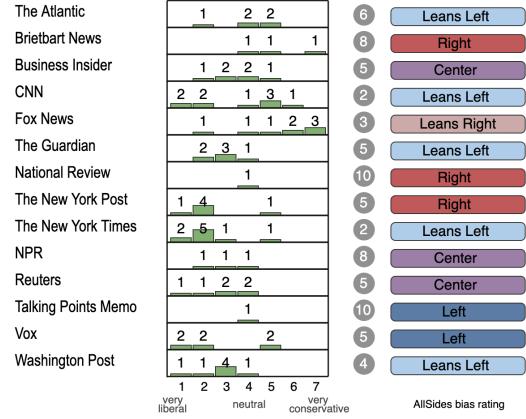


Figure 4: Before taking Study #2, novice news participants were asked to rate the perceived political biases of the news sites in the study’s data corpus. The histograms show the overall ratings for each site; gray circles indicate the number of participants who simply said “I don’t know.” This figure illustrates that Study #1 participants were largely unfamiliar with American news sites, and could therefore be considered novice news users. As a reference, the rightmost column shows each site’s media bias rating from AllSides [2].

$\sigma = 1.53$) and 8/11 participants agreed that they regularly kept up with American news stories, when we asked participants questions about specific news sites they were largely oblivious about how to rate news sites. Figure 4 shows this explicitly; the high variance in the histograms indicate uncertainty, and the gray circles to the right indicate the number of participants who simply answered “I don’t know” about a particular site. The labels in the right-most column indicate the site’s AllSides media watchdog rating, which likely represents the consensus rating for news experts. While bias is not a perfect proxy for the coverage diversity analysis that News Kaleidoscope supports, this chart effectively illustrates that the Study #2 participants were largely unfamiliar with the American news sites in our data corpus, and could therefore be considered news novices for the purposes of the study.

6.4 Study #2 Results

To analyze Study #2, we first briefly report high-level system ratings based on post-study questionnaire responses, and then analyze participant comments (collected both from think-aloud comments and post-study feedback) to qualitatively assess the types of insights and action patterns that News Kaleidoscope promotes for novice users.

6.4.1 System Ratings via Questionnaire Responses

Figure 5 shows questionnaire responses about News Kaleidoscope which describe overall system feedback and the perceived usefulness of its interface features. Overall system ratings (Q1–Q11) were generally positive, including that it was easy to learn, use, and comprehend (Q1–Q3), organized data into a high-level overview (Q5), supported meaningful analysis (Q6–Q9), and encouraged participants to think about coverage diversity in news reporting (Q10–Q11).

Responses about specific interfaces features (Q12–Q25) followed a similar trend in being generally well regarded. However, reviewing these responses individually indicates at least some features were not as well received, notably the word cloud in the NER panel (Q21–Q22) and the action of changing the number of clusters in the emotions panel (Q24).

Overall feedback on system

- Q1) Easy to learn
- Q2) Easy to use
- Q3) Easy to understand
- Q4) Retrieves relevant articles for a query
- Q5) Provides a good overview of a news event
- Q6) Easy to compare news sites for a specific event
- Q7) Shows high-level trends and patterns about a news event
- Q8) Organizes articles into meaningful groupings
- Q9) Provides sufficient information about why articles are grouped together
- Q10) Encourages me to think about how coverage can vary between sites
- Q11) Encourages me to think about potential biases contained in articles

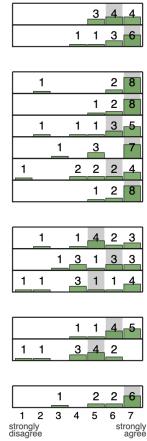
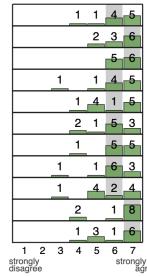


Figure 5: This chart shows overall ratings from the post-study questionnaire about various system aspects by the novice users in Study #2. Median ratings are indicated in grey.

6.4.2 Supporting Insights for Novice Users

To understand the insights of news novices, we reviewed both the think-aloud comments made during the exploration stage and the freeform comments from the review stage. Specifically, we contextualize these results to Study #2 to highlight the different experiences that News Kaleidoscope provides for non-domain experts.

For news novices, the interface was both easy to use and too complex. Despite not being the intended user base, we received several comments about the News Kaleidoscope being “useful” (p6), and “easy to use” (p7). However, some users paradoxically felt overwhelmed by the system’s available features and interactions. “*I didn’t use all the features as there were so many*” (p11). Multiple users suggested streamlining the user experience (p7, p6, p1), or thought that follow-up sessions would lead to more efficient analysis: “*Probably [the] user needs more time to make full use of the system*” (p5). This mixed feedback is likely due to a lack of domain knowledge by news novices, which hampered some of the users’ experiences while not being a problem for others.

Aspects of the user experience echoed news experts. Similar to Study #1, several participants remarked how News Kaleidoscope allows them to review articles in a scalable fashion: “*I can generally just read one article from one site on my news app or site, however this system is helping me read a number of articles on an event by different sites at the same time. This in addition to the other features such as keywords, time, entities and emotions is really helpful*” (p8).

Entities, both via tooltips and in the NER panel, were especially helpful for getting a quick overview of a story: “*Entities tell me who and what all are included without reading the entire article*” (p6). Some users leveraged the entity-based pairwise similarity in the NER panel’s adjacency matrix, as this was the only visualization that directly compared two articles: “*There is more similarity in*

news articles on people than other two” (p3).

Clustering is effective for showing coverage diversity. Several participants described the clustering—both in the ordination and emotions panels—as an effective and “intuitive” (p7) way to show coverage diversity. “*When I look at the clusters [in the MDS plot], there is a shift in polarity from right to left. Here towards the right there are more articles on Hillary and here’s there’s more Trump*” (p11). In the MDS plot, clustering “*helped to select which articles to read*” (p8) and make subselections on.

In the emotions panel, clustering “*allowed me to get a good idea of emotional values from the news articles*” (p6). Several participants mentioned the emotional style vectors provided helpful insight into how and why articles were clustered, thus making the emotions panel “*better than clustering based on keywords and entities as it was more intuitive than keyword clustering*” (p6).

At times however, there was confusion when computed clusters did not line up with participant expectations or mental models. For example, in the ordination panel participants would sometimes wonder why a particular story was included in a particular cluster, and in the emotions panel participants sometimes were unsure how emotions contributed to a cluster. As an example, when p2 was reviewing articles about the Pulse Nightclub shooting: “*It’s all fear and anger and surprisingly there is no sadness which one would normally expect. I didn’t expect that*.”

Participants recognized diversity, but it was hard to translate these into broader conclusions about media bias. Along these lines, while participants were able to recognize variation between articles and sites, they were reluctant to use these to draw explicit conclusions about media bias. As most users were unfamiliar about the perceived political leanings of media sites, many were unwilling to make strong statements about specific sites or bias based on keyword, entity, and emotional variations in the stories. Statements tended to focus on the specific articles that were being analyzed: “*Not sure if CNN is anti- or pro-[Trump], but its articles show that they are doubting Trump’s win*” (p7).

Interestingly, some participants contextualized news sites using a prior they were familiar with: clickbait writing. “*Atlantic has clickbait headlines*” (p2). “*CNN kind of publishes reactionary kind of headlines but Reuters has a more balanced view*” (p4). “*This site has clickbait articles, the headlines would make you think something but the emotions in the tooltip suggest something else*” (p2). Several participants also suggested the system explicitly anchor or label the bias and framing of articles or news sites to help contextualize how they should be interpreted. “*Naming the clusters would have made user investigation faster*” (p6). “*I see how on an event is covered, but I couldn’t form an opinion [about specific news sites] . . . so if more annotations were present it would have been better*” (p5).

6.5 Study #3

Studies #1 and #2 validated several aspects of News Kaleidoscope, particularly its ability to support coverage diversity for news experts (as well as its ability to introduce the topic to news novices). Despite the largely positive feedback, some features were commented upon as either missing or confusing by multiple study participants. Specifically, we identified three targeted system improvements based on participant feedback: (1) provide additional transparency or explanation into the number of clusters to show, both in the ordination and emotions panels, (2) provide views that support summary and comparative analysis of news events and sites, and (3) provide additional explanation of how emotions are computed. The systems features in Figure 2(b5, b6, b7, d5) were thus added to the system.

To validate these design improvements, we followed up with four participants (u1–u4) from Studies #1 and #2 (three from Study #1, one from Study #2), who could compare the system’s updates against the original version. For each participant, we conducted a pair analytics session over Zoom, demoing the system’s new functionality and soliciting freeform feedback. In general, the new components

were positively appreciated. “*I like the site overview part because, the first time I used the system, it felt like a lot of information was thrown at me and I didn’t know where to start from*” (u1). “*The cluster annotation is really good. Earlier it was a little harder than the present system*” (u3). All three expert users (u1–u3) reiterated their earlier comments that systems like News Kaleidoscope would be useful for their work, highlighting the lack of accessible visual analytics systems for studying news corpuses: “*I still continue to think this will be a useful tool. It’s something I could see myself using in research.*”(u3).

7 DISCUSSION

While there are many existing text visualization tools in existence, based on our experience in developing and evaluating News Kaleidoscope, we believe that interfaces tailored specifically for news reporting analysis can better benefit both for news experts and news novices. News Kaleidoscope design was intended to elicit nuanced analysis of coverage diversity via the subselection workflow. To understand the implications of this design study, we compare the insights and workflow themes between the news expert and novice users, and discuss how they can guide the design of visualization systems for news reporting analysis.

Novice news users recognize diversity, news experts can see research possibilities. While both user groups agreed that News Kaleidoscope supports the analysis of coverage diversity, one main takeaway from the studies is that insights differ based on the user’s experience. Novice news users were able to assert that coverage diversity exists in the articles written about news events, but generally could not make strong conclusions or broad inferences about bias. News experts, already familiar with coverage diversity and media bias, were able to use the system to see deeper nuances about coverage diversity. Analysis systems for news novices can be tailored to accommodate their lack of domain expertise.

Explainability on demand improves sensemaking. No users in our evaluation were experts in data visualization or NLP. While both user groups could understand and interact with the interface’s visualizations, many participants wanted more explainability, particularly when data points or clusterings displayed in unintuitive or unexpected ways. Such desires echo recent trends in explainable machine learning and artificial intelligence, where transparent and interpretable models are desired to promote trust in and understanding of predictions [30]. The explainability features added for Study #3 were non-disruptive to the overall user experience, but they were positively received in our follow-up sessions with participants.

The need for visualization tools in journalism research. Study #1 participants were researchers in journalism and political science, but none had previously used visual analytics tools like this. While systems similar to News Kaleidoscope have previously been published in the visualization community, the lack of adoption by our participants indicates that wider dissemination would benefit other communities. Tools like this also motivate an interesting question for these communities: *If coverage diversity is identified, how can we reduce media bias?* This is a non-trivial problem, but we believe visualization can likely provide an important step in the process, by helping researchers identify where issues are present.

Visually analyzing the temporal dynamics of news coverage. Some participants in Study #1 noted temporal analysis as an important facet of coverage diversity: “*Development of news stories over time is interesting ... I personally for my research loved the time evolution in both the cases especially the emotions over time, that is a really informative information in journalism*” (e7). While News Kaleidoscope provides a limited amount of temporal analysis, systems that focus on the temporal (and unique) dynamics of news stories are lacking. For example, reporting about events evolves over time as more facts are learned and media analysis and commentary is conducted. Future visual analytics systems can be tailored towards

this type of analysis, and we intend to extend News Kaleidoscope to better focus on this type of evolutionary analysis.

Explicitly labeling bias in news stories? Several news novices requested explicit bias labeling of articles and news sites, however we are cautious that such an approach is necessary or even appropriate. *Ex ante* labeling inherently reduces the power of empirical analysis by anchoring users to the classification outputs of a model. When considering coverage diversity in the context of specific news events, explicit bias labeling might poorly account for sites that publish contrasting viewpoints such as editorials and opinion articles. Interfaces that can provide an initial bias labeling while accounting for such nuances (perhaps via uncertainty or probabilistic techniques) are one strategy, but we leave this as future work.

Alternate visual encodings for News Kaleidoscope Though we found that News Kaleidoscope successfully supports the analysis of the coverage diversity of news articles, in the future we intend to look at ways to improve the analytic process via alternative visualization and analytical approaches. As an example, instead of showing a flat clustering of retrieved articles (in the Ordination Panel), a hierarchical topics tree could instead be constructed, with individual nodes representing topics constructed from relevant keywords, entities, and/or biases. Coverage diversity could be demonstrated by analyzing and interacting with the tree’s structure; for large trees, aggregation and simplification techniques would likely need to be employed (e.g., [49]).

8 CONCLUSION

We contribute News Kaleidoscope, an interactive visual analytics system that supports analysis of news events with a focus on the coverage diversity of reporting articles. News Kaleidoscope is designed based on a formal task abstraction for news experts and combines several NLP and visualization techniques into an accessible user experience based around keyword subselections in article sets. Based on a holistic, three-part evaluation, we find that News Kaleidoscope supports different types of insights based on the journalistic domain expertise of the user: news novices are able to recognize coverage diversity at a high-level, but news experts can contextualize it for deeper insight, such as by characterizing articles according to their extracted emotional styles. Study results motivate several guidelines and takeaways for future systems that visualize coverage diversity, including the importance of explainability for non-technical users and that the temporal, comparative, and summary dynamics of news events and coverage can be emphasized.

REFERENCES

- [1] Trends and facts on online news — state of the news media — pew research center. <https://www.journalism.org/fact-sheet/digital-news/>.
- [2] Media bias ratings — allsides. <https://www.allsides.com/media-bias/media-bias-ratings>, 2020.
- [3] The Center for Media and Public Affairs — George Mason University. <https://cmqa.gmu.edu/>, 2020.
- [4] Types of media bias and how to spot it — allsides. <https://www.allsides.com/media-bias/how-to-spot-types-of-media-bias>, 2020.
- [5] E. Alexander, J. Kohlmann, R. Valenza, M. Witmore, and M. Gleicher. Serendip: Topic model-driven visual exploration of text corpora. In *2014 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 173–182, IEEE, 2014.
- [6] R. Arias-Hernandez, L. T. Kaastra, T. M. Green, and B. Fisher. Pair analytics: Capturing reasoning processes in collaborative visual analytics. In *2011 44th Hawaii international conference on system sciences*, pp. 1–10. IEEE, 2011.
- [7] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022, 2003.
- [8] N. Cao and W. Cui. Overview of text visualization techniques. In *Introduction to Text Visualization*, pp. 11–40. Springer, 2016.

- [9] Y. Chen, L. Wang, M. Dong, and J. Hua. Exemplar-based visualization of large document corpus. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):1161–1168, 2009.
- [10] H. L. Chieu and Y. K. Lee. Query based event extraction along a timeline. In *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 425–432, 2004.
- [11] C. Collins, G. Penn, and S. Carpendale. Bubble sets: Revealing set relations with isocontours over existing visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):1009–1016, 2009.
- [12] W. Cui, S. Liu, Z. Wu, and H. Wei. How hierarchical topics evolve in large text corpora. *IEEE transactions on visualization and computer graphics*, 20(12):2281–2290, 2014.
- [13] G. Da San Martino, S. Shaar, Y. Zhang, S. Yu, A. Barrón-Cedeño, and P. Nakov. Prta: A system to support the analysis of propaganda techniques in the news. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pp. 287–293. Association for Computational Linguistics, Online, July 2020. doi: 10.18653/v1/2020.acl-demos.32
- [14] S. DellaVigna and E. Kaplan. The fox news effect: Media bias and voting. *The Quarterly Journal of Economics*, 122(3):1187–1234, 2007.
- [15] N. Diakopoulos, D. Elgesem, A. Salway, A. Zhang, and K. Hofland. Compare clouds: Visualizing text corpora to compare media frames. In *Proceedings of IUI Workshop on Visual Text Analytics*, pp. 193–202, 2015.
- [16] N. Diakopoulos, A. X. Zhang, and A. Salway. Visual analytics of media frames in online news and blogs. In *Proc. IEEE InfoVis Workshop on Text Visualization*, 2013.
- [17] W. Dou, X. Wang, R. Chang, and W. Ribarsky. Paralleltopics: A probabilistic approach to exploring document collections. In *2011 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 231–240. IEEE, 2011.
- [18] W. Dou, X. Wang, D. Skau, W. Ribarsky, and M. X. Zhou. Leadline: Interactive visual analysis of text data through event identification and exploration. In *2012 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 93–102. IEEE, 2012.
- [19] N. Elmquist and J. S. Yi. Patterns for visualization evaluation. *Information Visualization*, 14(3):250–269, 2015.
- [20] J. R. Finkel, T. Grenager, and C. Manning. Incorporating non-local information into information extraction systems by gibbs sampling. In *Proceedings of the 43rd annual meeting on association for computational linguistics*, pp. 363–370. Association for Computational Linguistics, 2005.
- [21] M. E. Fonteyn, B. Kuipers, and S. J. Grobe. A description of think aloud method and protocol analysis. *Qualitative health research*, 3(4):430–441, 1993.
- [22] C. Glynn, S. Herbst, M. Lindeman, and G. J. O’Keefe. *Public Opinion*. Routledge, 2015.
- [23] T. Groseclose and J. Milyo. A measure of media bias. *The Quarterly Journal of Economics*, 120(4):1191–1237, 2005.
- [24] F. Hamborg, K. Donnay, and B. Gipp. Automated identification of media bias in news articles: an interdisciplinary literature review. *International Journal on Digital Libraries*, 20(4):391–415, 2019.
- [25] E. Humprecht and F. Büchel. More of the same or marketplace of opinions? a cross-national comparison of diversity in online news reporting. *The International Journal of Press/Politics*, 18(4):436–461, 2013.
- [26] M. Kim, K. Kang, D. Park, J. Choo, and N. Elmquist. Topiclens: Efficient multi-level visual topic exploration of large-scale document collections. *IEEE transactions on visualization and computer graphics*, 23(1):151–160, 2016.
- [27] J. B. Kruskal and M. Wish. *Multidimensional scaling*, vol. 11. Sage, 1978.
- [28] K. Kucher, C. Paradis, and A. Kerren. The state of the art in sentiment visualization. In *Computer Graphics Forum*, vol. 37, pp. 71–96. Wiley Online Library, 2018.
- [29] H. Lee, J. Kihm, J. Choo, J. Stasko, and H. Park. ivisclustering: An interactive visual document clustering via topic modeling. In *Computer graphics forum*, vol. 31, pp. 1155–1164. Wiley Online Library, 2012.
- [30] Z. C. Lipton. The mythos of model interpretability. *Queue*, 16(3):31–57, 2018.
- [31] S. Liu, X. Wang, C. Collins, W. Dou, F. Ouyang, M. El-Assady, L. Jiang, and D. A. Keim. Bridging text visualization and mining: A task-driven survey. *IEEE transactions on visualization and computer graphics*, 25(7):2482–2504, 2018.
- [32] E. Loper and S. Bird. Nltk: The natural language toolkit. In *Proceedings of the ACL Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics*. Philadelphia: Association for Computational Linguistics, 2002.
- [33] D. Luo, J. Yang, M. Krstajic, W. Ribarsky, and D. Keim. Eventriver: Visually exploring text collections with temporal references. *IEEE transactions on visualization and computer graphics*, 18(1):93–105, 2010.
- [34] L. v. d. Maaten and G. Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(Nov):2579–2605, 2008.
- [35] J. MacQueen et al. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, vol. 1, pp. 281–297. Oakland, CA, USA, 1967.
- [36] C. D. Manning, M. Surdeanu, J. Bauer, J. R. Finkel, S. Bethard, and D. McClosky. The stanford corenlp natural language processing toolkit. In *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations*, pp. 55–60, 2014.
- [37] L. McInnes, J. Healy, and J. Melville. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*, 2018.
- [38] S. M. Mohammad and P. D. Turney. Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon. In *Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text*, pp. 26–34. Association for Computational Linguistics, 2010.
- [39] R. Plutchik. The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American scientist*, 89(4):344–350, 2001.
- [40] M. Pollard and J. Kavanagh. *Profiles of News Consumption: Platform Choices, Perceptions of Reliability, and Partisanship*. RAND Corporation, 2019.
- [41] J. Ramos et al. Using tf-idf to determine word relevance in document queries. In *Proceedings of the first instructional conference on machine learning*, vol. 242, pp. 133–142. Piscataway, NJ, 2003.
- [42] P. J. Rousseeuw. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65, 1987.
- [43] M. Sedlmair, M. Meyer, and T. Munzner. Design study methodology: Reflections from the trenches and the stacks. *IEEE transactions on visualization and computer graphics*, 18(12):2431–2440, 2012.
- [44] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings 1996 IEEE symposium on visual languages*, pp. 336–343. IEEE, 1996.
- [45] J. Stasko, C. Görg, and Z. Liu. Jigsaw: supporting investigative analysis through interactive visualization. *Information visualization*, 7(2):118–132, 2008.
- [46] R. Swan and D. Jensen. Timelines: Constructing timelines with statistical models of word usage. In *KDD-2000 Workshop on Text Mining*, pp. 73–80. Citeseer, 2000.
- [47] A. Thompson. All the news: 143,000 articles from 15 American publications. <https://www.kaggle.com/snapcrack/all-the-news>, 2017.
- [48] S. Tilkov and S. Vinoski. Node.js: Using javascript to build high-performance network programs. *IEEE Internet Computing*, 14(6):80–83, 2010.
- [49] K. Wongsuphasawat, D. Smilkov, J. Wexler, J. Wilson, D. Mane, D. Fritz, D. Krishnan, F. B. Viégas, and M. Wattenberg. Visualizing dataflow graphs of deep learning models in tensorflow. *IEEE transactions on visualization and computer graphics*, 24(1):1–12, 2017.
- [50] C. C. Yang, X. Shi, and C.-P. Wei. Discovering event evolution graphs from news corpora. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 39(4):850–863, 2009.