Exploring Evolving Media Discourse Through Event Cueing

Yafeng Lu, Michael Steptoe, Sarah Burke, Hong Wang, Jiun-Yi Tsai,
Hasan Davulcu, Douglas Montgomery, Steven R. Corman, Ross Maciejewski, *Senior Member, IEEE*

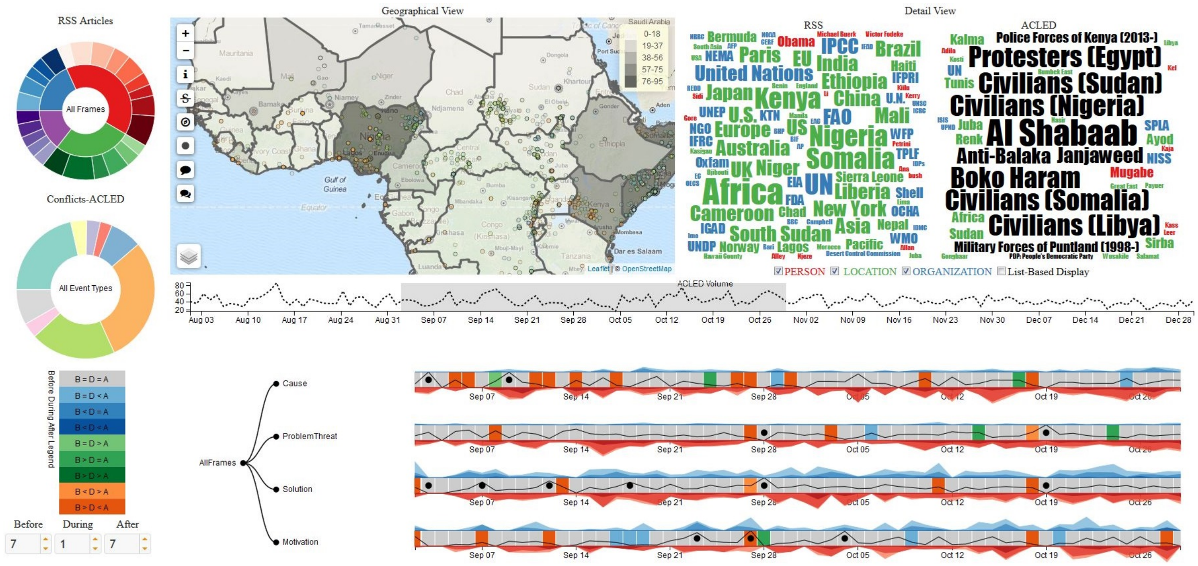


Fig. 1: Overview of the event cueing visual analytics framework. The map view provides a geographical visual analytics environment to enable exploration of frames and entities over space and time. The detailed view to the right of the map switches between entity wordles and list-based displays. The time series view contains a hierarchical frame analysis visualization. Each line visualizes significant events and the sentiment associated with a media frame or a frame class in the expanded leaf nodes. The control pane, which consists of the top left donuts, shows the distribution of frames and events and is used to filter categorical variables in the linked views.

Abstract— Online news, microblogs and other media documents all contain valuable insight regarding events and responses to events. Underlying these documents is the concept of framing, a process in which communicators act (consciously or unconsciously) to construct a point of view that encourages facts to be interpreted by others in a particular manner. As media discourse evolves, how topics and documents are framed can undergo change, shifting the discussion to different viewpoints or rhetoric. What causes these shifts can be difficult to determine directly; however, by linking secondary datasets and enabling visual exploration, we can enhance the hypothesis generation process. In this paper, we present a visual analytics framework for event cueing using media data. As discourse develops over time, our framework applies a time series intervention model which tests to see if the level of framing is different before or after a given date. If the model indicates that the times before and after are statistically significantly different, this cues an analyst to explore related datasets to help enhance their understanding of what (if any) events may have triggered these changes in discourse. Our framework consists of entity extraction and sentiment analysis as lenses for data exploration and uses two different models for intervention analysis. To demonstrate the usage of our framework, we present a case study on exploring potential relationships between climate change framing and conflicts in Africa.

Index Terms—Media Analysis, Time Series Analysis, Event Detection

1 Introduction

Recently, the visual analytics community has begun developing a variety of tools for analyzing media collections. These tools tend to focus on event detection from text streams [39], correlation analysis [28], and topic evolution [15]. These tools are often concerned with understanding an ongoing narrative from structured text and focus on enabling the user to place news stories within the context of other ongoing events. However, very few tools [11, 12, 13] explore how media is being framed, and, to our knowledge, none have explored changes in frames over time and space. In studying public communications, framing is the use of rhetorical devices (e.g., words, phrases, metaphors, images) to encourage one interpretation of a set of facts and discourage other interpretations [16]. Examples include efforts by U.S. conservatives in the 1990s to reframe the estate tax as a "death tax", and competing frames of the Occupy Wall Street protests, "the 99%" (vs. the 1%) as opposed to "makers vs. takers". Framing affects the attitudes and behavior of audiences [9], and it is also regarded as a key media effect, in that media "actively set the frames of reference that readers or viewers use to interpret and discuss public events" [35].

• *Yafeng Lu, Michael Steptoe, Sarah Burke, Hong Wang, Jiun-Yi Tsai, Hasan Davulcu, Douglas Montgomery, Steven R. Corman, and Ross Maciejewski, are with Arizona State University. E-mail: {lyafeng,msteptoe,seburke2,hxwang,jtsai8,HasanDavulcu, doug.montgomery,steve.corman, rmacieje}@asu.edu.*

Manuscript received 31 Mar. 2015; accepted 1 Aug. 2015; date of publication xx Aug. 2015; date of current version 25 Oct. 2015. For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org.

Understanding framing in the media is vital as it influences the way people interpret the topic under analysis. Framing is also critical to the success of social movements and can be a driver for change or stagnation [4]. What is of interest is how these frames are applied and how they evolve over time in the context of other events. However, it can be quite difficult to determine when changes in framing occur and what events may have contributed to changing attitudes.

In this work, we present a visual analytics framework for event cueing from media. For a given collection of documents (related by topic and coded with frames), we enable analysts to explore ongoing media discourse with respect to the overall framing and related sentiment of the narratives. In order to understand when and how framing about a topic has shifted, we employ intervention models for time series analysis. Such models examine how a measure changes over time and how this measure is affected by some external event, or intervention, at a given time t. If the measure is significantly different before and after the intervention, then one can hypothesize that an intervention is associated with a change in the measurement. By highlighting these statistically significant intervention points, we can cue analysts to time periods of interest. Then, by linking the media data source with secondary sources of information relevant to the topic, an analyst can explore the frame evolution within the context of ongoing events.

This work is directly related to previous works, such as Narratives [18] and EventRiver [27], which focus on placing media stories into their historical and social context by allowing people to explore topics and keywords and associate them with other ongoing stories and events. Unlike previous work, our framework utilizes intervention modeling strategies and multisource data. Our goal is to enable analysts to cue to important dates in the dataset. Media can then be explored in the context of the changing sentiment of the framed documents as well as linked to concurrent events that may have impacted the media discourse. While previous work from Diakopolous et al. [11, 12, 13] developed tools for frame analysis, their work provided no support for entity extraction, sentiment analysis, or linking multisource data. Our contributions include:

1) An ensemble of intervention modeling techniques for event cueing and hypothesis generation,

2) The application of visual analytics for media framing in the context of entities, sentiment, geography and multisource data.

2 Related Work

As media sources have broadened from network news coverage to microblogs, Twitter, etc., a variety of tools and techniques have been developed to analyze and explore such data sources. Given that media data generates events over time in unstructured text, the majority of tools and techniques developed have focused on temporal visualizations, topic analysis, and pattern and anomaly detection.

2.1 Time Series Visual Analytics

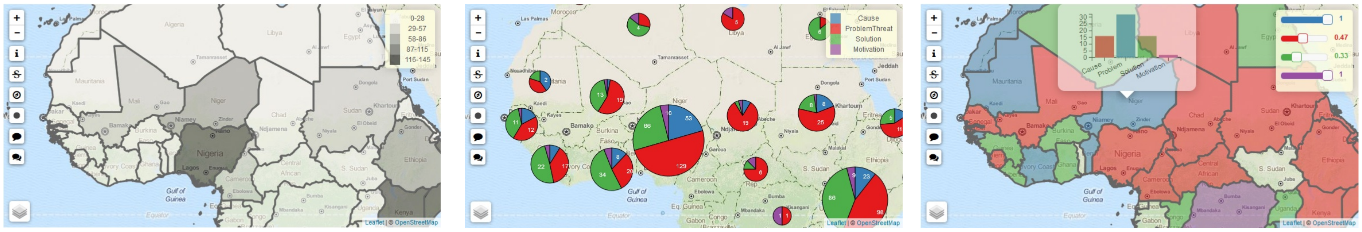
Visualization has been successfully applied to analyze time-oriented data, most commonly through the use of line graphs and their variations [17], as well as calendar views and clock views for periodical or seasonal patterns [3, 22, 37]. A variety of enhancements to these techniques have been proposed over the years to enable better sensemaking of events and records. For example, LifeFlow [42] combines a list-based display for intra-record pattern analysis and an aggregated overview display for inter-record trends analysis to visualize time-point based event sequences. EventFlow [30] extends LifeFlow to handle interval events and explore the relationship between event sequences and associated outcomes. Another extension of LifeFlow, Outflow [41] aggregates multiple event sequences, visualizes them as the pathways through different event states, and connects the pathways to their associated outcomes so that users can explore progression paths and results. Of interest to our work is that OutFlow also incorporates external factors which may influence the event sequence. Our work differs in that we focus on cueing analysts to events in time series datasets through the use of intervention models. These models enable users to find sequences in the data that appear to have deviated. Our framework then links these deviations to external data sources to identify potential causes to these deviations.

The incorporation of statistical techniques into time series visualization has led to the development of a variety of visual analytics solutions. A typical example is the visual analytics process proposed in Bogl et al. [5] where visualization is used to guide domain experts in statistical model parameter selection. Their prototype system, TiMoVA, is developed to facilitate the process of parameter settings in autoregressive integrated moving average (ARIMA) and seasonal ARIMA models. A probabilistic decision tree learner is used in the e-transaction time-series visual analytics system VAET in [43] to explore transaction patterns among multiple users in a temporal context. These tools focus on enabling users to visually develop statistical models of the data. In contrast, our work focuses primarily on using statistical models for cueing analysts to events of interest in the time series.

Our proposed type of cueing is similar to work in event detection, and visual analytics has posed a variety of solutions for anomaly and event detection [8]. Classification-based event detection methods have been applied in many visualization systems. For example, Scatterblogs [6, 36] is a scalable system enabling analysts to find quantitative information and detect spatiotemporal anomalies within a large volume of geo-located microblog messages. Work by Chae et al. [7] utilizes a seasonal-trend-decomposition method to determine anomalous changes in topics in social media. Gotz et al. developed DecisionFlow [20], which integrates interactive multi-view visualizations and ad hoc statistical models to support the analysis of high-dimensional temporal event sequence data. While a variety of statistical methods have been applied for visual analytics of temporal data, these methods typically focus on anomalous behavior. In our framework, we focus on the concept of an external intervention causing the system to deviate. This framework requires different statistical analysis and also needs to integrate multi-source data for analysis. To our knowledge, this approach is the first such application in visual analytics to explore time series data in the context of interventions.

2.2 Media Visual Analytics

While applicable to a variety of domains, our focus is specifically on media data, such as online news and microblogs. Recently, much attention has been paid to this domain area in the visual analytics community, with techniques focusing on knowledge expression, topic extraction, pattern analysis, and storytelling [14, 19, 21, 23, 24]. CloudLines [23] focuses on the detection of visual clusters in a compressed view of multiple time series to enable the scalable analysis of media streams. To improve sensemaking, LeadLine [14] explores named entities, locations, and bursts of topic related events by visualizing the shift of topic volume for different time streams and emphasizing detected events. Contextifier [21] is designed for contextualizing visualization by providing customized annotations for the stock timeline graph with reference to the content in a news article. StoryTracker [24] combines interactive visualization and text mining techniques to facilitate the analysis of similar topics that split and merge over time. NewsViews [19] is a novel automated news visualization system that creates thematic maps automatically for news articles. It leverages text mining to identify key concepts and locations discussed in articles. TopicPanorama [25] visualizes the full picture of relevant topics from different sources to analyze common and distinctive topics. Similar to previous works, we also leverage text mining techniques on media articles. Our system extracts entities and their associated geolocations. Unlike previous works on media frame visual analytics [11, 12, 13], which typically focus on topic analysis and cooccurring words, our system focuses on frame analysis in conjunction with multisource data. We focus on a single topic and explore how it is being discussed (i.e., framed), rather than focusing on multiple topics. In this work, frames are organized into a hierarchical set and the change in how documents are framed (with respect to space and time) can be explored. By visualizing statistical results together with the hierarchical frames, we can enhance the hypothesis generation process.



(a) Choropleth (b) Cluster Pie (c) Weighted Category Choropleth

Fig. 2: Categorical data spatial distribution visualization view. View (a) shows the default choropleth map which colors each country based on the density of all frames. View (b) shows pie glyphs on the map displaying the proportional distribution of different frame categories in each cluster. View (c) shows a weighted choropleth map which colors each country based on the weighted frame density. The weights on each category can be changed interactively by the analysts.

3 Event Cueing Environment

The goal of our visual analytics framework is to facilitate the hypothesis generation process by linking multisource data through statistical event cueing in the form of intervention models. Our framework consists of three main views: 1) the spatial view (top left Figure 1), which visualizes the geographic location of media steams and events coded in secondary data sources; 2) the detail view (top right Figure 1), which provides a lens into the media text and detailed descriptions of events from secondary data sources chosen by the analyst, and; 3) the time series view (bottom Figure 1), which shows a hierarchical frame-coded, time-orientated media stream with sentiment and intervention analysis. All views are linked by the overview timeline shown in the middle of Figure 1 which displays the trend of a secondary dataset.

3.1 Task Characterization

The basis for this work is founded on an interdisciplinary collaboration between computer science and communication. Partners from the Hugh Downs School of Human Communication at Arizona State University are interested in applying their knowledge of framing to issues of national security risks related to climate change. Their work focuses on exploring the framing of climate change research in Africa and how (if at all) this is impacted by ongoing conflicts in the region. They posit that, in order to understand how the media reflect different organizations' interests in addressing climate change as a social problem, it is necessary to supplement the social movement focus on resource mobilization to framing processes of collective action problems. To do this, they developed a nuanced typology for studying climate change framing and its adequacy for supporting social movement that would be necessary to overcome the collective action problem. They apply this framework to examine framing of climate change in media and social media texts collected from the Niger Basin region over eight months from August 2014 to March 2015, applying a novel coding technique to assess diagnostic, prognostic and motivational framing as the keys to effective social movements. While the datasets and examples given in this application focus on media with regards to climate change and social unrest, our techniques can be adopted to any multi-source data in which analysts are looking for changes in media frames due to associated events (for example, severe flooding, prolonged droughts). We have identified three major intents of the communication scientists in the context of media analysis:

1) Analysts would like to know how frames are spatially distributed to understand the international context of framing;

2) Analysts would like to know when the distribution of frames change and quickly be able to explore events that may have impacted this change in media framing;

3) Analysts would like to know what people, locations and organizations are being discussed in the media before, during and after changes in framing occur.

As such, our framework has been designed to support the spatiotemporal analysis of frames and cue analysts to when the distribution of frames has changed. These cues then suggest time windows in which to explore links to secondary datasets.

3.2 Datasets

To illustrate our framework, we use a climate change media dataset and the Armed Conflict Location & Event Data Project (ACLED) dataset [1] as an example. However the proposed framework is flexible for analyzing any media data.

Media: The media dataset is composed of RSS feeds from 122 English language news outlets in the Niger basin countries since August 2014. RSS feeds were scanned hourly and filtered for relevance in a two-stage process. First, news texts were matched against a set of 222 keywords developed from the Intergovernmental Panel on Climate Change (IPCC) report and supplemented by project experts. Subsequently, texts passing the keyword test were analyzed by a machine classifier, trained on a set of 1,000 texts classified by coders as relevant or irrelevant to social discourse of climate change. News articles passing both tests were placed into the database for analysis. The RSS news dataset collected 1245 relevant articles with 9070 sentences. For this study, each sentence was coded by trained coders into one (or none) of 25 categories comprising four classes (cause, problems/threat, solution, motivation) that represent different types of framing for climate change. Then each article was represented by a vector of frame counts normalized by the number of sentences coded. The average Krippendorff α reliability of the coders on a set of training documents was 0.81 and judged to be acceptable. Future work will use trained classifiers for frame extraction.

ACLED: The ACLED dataset contains information on the dates and locations of all reposted political violence events in over 50 developing countries, with a focus on Africa. Each event record contains information on date, location, event type and actors involved. From August to December 2014, it contains approximately 6500 events.

3.3 Media Data Processing

Media messages contain large amounts of information which can be complex to effectively analyze. Our framework applies a variety of automatic data preprocessing techniques including entity, geolocation, and sentiment extraction.

Entity Extraction: Entities, such as person names, locations, and organizations, inform much of the underlying media discourse. A variety of named entity recognition methods have been proposed for different contexts in natural language processing. We applied the well-known natural language processing tool CoreNLP [29] to a streaming RSS news dataset and the secondary dataset (ACLED in our example) to extract named entities. For the 1245 articles from August to December 2014, we have 19,756 entities in which there are 2107 persons, 5791 locations and 3146 organizations extracted from the RSS dataset. The same entity recognition process was performed on the ACLED dataset for the notes in each record, extracting 367 persons, 998 locations, 286 organizations.

Geolocation Extraction: An article may have location attributes, either based on where the article is posted or the region the article

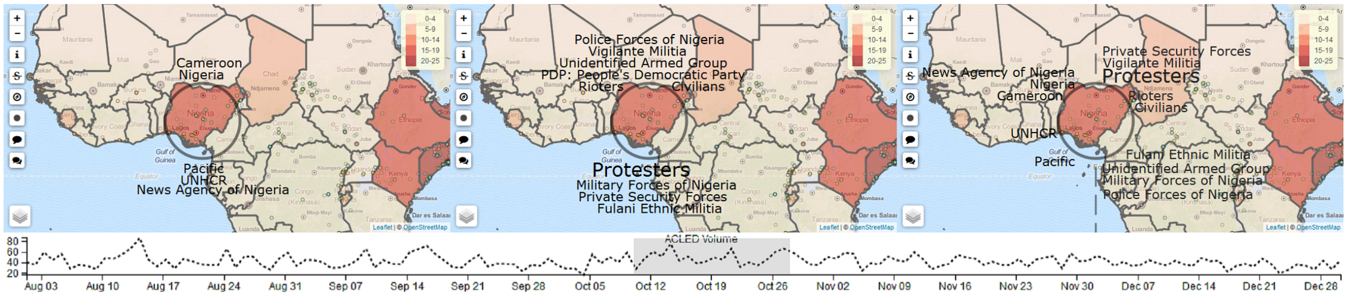


Fig. 3: Entity lens on the map shows the most frequently appearing entities recognized from documents that are geo-located in the lens' area. The left figure shows the named entities for the RSS news dataset, the middle figure shows the actors given in the ACLED dataset,and the right shows a comparison lens with the RSS news' entities on the left and ACLED events' actors on the right. The example shown in this Figure covers the data from Oct. 11th to Oct. 27th for all the ACLED event type and problem frames in the RSS news.

discusses; however, this information is not always explicitly coded. Given that framing may differ by geographic region, our framework preprocesses the media stream to extract geographic locations. We use the Data Science Toolkit [2] to extract and geocode this information.

Sentiment Extraction: Media data encapsulates information about events, responses, and reviews. In exploring media data, sentiment analysis can provide a quick overview of the attitude a media document's author might have with regards to the underlying story. To extract the sentiment embedded in the media data, three sentiment analysis classifiers are applied at the per sentence level. Details on the classifiers are provided in Section 3.6.

3.4 Geographical View

Both media data and event sequences from the secondary dataset have geolocation information. The geographical view is built to explore the distribution of frames and compare entities between media data and other data sources.

3.4.1 Exploring Spatial Distribution of Frames

Previous work on frame visualization focused on document keywords. In this work, we want to allow users to explore frames by country, entities, and sentiment. To analyze the spatial distribution of different frames, we created a categorical spatial data distribution visualization view, Figure 2. To show the cumulative frame distribution of a dataset, Figure 2(a) displays a choropleth map colored by the density of frames in each country. Users can select any subset of frame categories to analyze. If only one class of frames are selected, the map color matches the color of the class, otherwise it uses gray. A drawback of this visualization is that only one variable/feature of the underlying data can be represented, even though there are multiple categories of frames in the data. To allow for multivariate encoding, we also use a symbol map with a pie chart, where each segment of the pie represents the number of sentences of a given frame (Figure 2(b)), and a weighted category choropleth map (Figure 2(c)) where colors correspond to a multivariate criteria function. Additionally, a tooltip displaying the histogram of different categories within a country is enabled to help better explore frame distributions.

3.4.2 Exploring Geo-located Entities

Our framework considers two types of entities in the data. One is recognized name entities, which are people, locations and organizations. The other is the predefined entities that may exist in the structured datasets that an analyst wishes to explore in the context of media discourse. We created an entity lens to explore geo-located text data. The geocoding of the entities derives from a sentence's geocoding for the RSS news dataset and the reported geolocation from the ACLED dataset. The entity lens is shown in Figure 3, where the most frequently referenced entities within the lens's area are extracted and organized around the lens. The most frequent entities are mapped closest to the lens's circumference based on available canvas space. The font size is dependent on the entities' frequency within the lens's circumference. The more frequent an entity is, the larger the text.

To link different datasets and find relationships between them, this entity lens has three modes: media data entity mode which shows only the RSS news entities (Figure 3(Left)), secondary data entity mode which shows only the ACLED actors (Figure 3(Middle)), and the combination mode which is a two-sided lens to encode entities for multi-source data (Figure 3(Right)). The combination mode shows the top entities from the RSS news dataset on the left of the lens and ACLED actors on the right with a dashed line in the middle to separate one from the other. All modes are also enabled in a coordinated view in which the lens can move over the map and update the wordles.

3.5 Hierarchical Frames Timeline View

The previous views are necessary to allow overview and detail views; however, the major contribution of this paper is the event cueing which is enabled through the hierarchical frames timeline. Previously mentioned techniques enable exploratory data analysis, the problem is that purely exploratory techniques put the burden of analysis completely on the analyst. Our goal is to cue the analyst to events that are statistically interesting in the data. To enable this, we begin with the timeline view showing the relative volume of frames per day. Specifically, each document has a number of sentences that are encoded with a single frame. The percent of framing of a document is the number of sentences in a document associated with a given frame divided by the total number of sentences framed. The frame volume can be visualized by the average document percent per day, the average number of sentences encoded with a frame across all documents in a day, or a variety of other metrics.

To detect possible interventions, we applied two time series analysis models and visualized the results on the timeline to cue analysts' exploration. In addition, sentiment information associated to the underlying data is also revealed by a two-side uncertainty-based stack river. Figure 4 shows our hierarchical timeline view of the media frames. Here an analyst expanded the cause frame to explore sentences that frame the cause of climate change to be due to human, natural effects, policies, or one uncertain of the cause.

Our approach is centered around the concept of an intervention. We assume that there may be events intervening with media reactions that cause a shift in how frames are distributed. We use statistical hypothesis test to detect the intervention dates. For each day in the dataset, we assume an intervention may have occurred. If a date under test indicates that the times before and after are statistically different, this then cues an analyst to explore related datasets to help enhance their understanding of what (if any) events may have triggered these changes in discourse.

Intervention Modeling Intervention models are used to explore what (if any) impact there is between an event and some secondary measure, for example, the impact that 9-11 had on George Bush's approval rating. In this case, we consider our events to be armed conflicts and the measure to be the amount of sentences that are framed in a document with respect to one of the 25 climate categories. Note that such models can be used for any event and measure dataset combination.

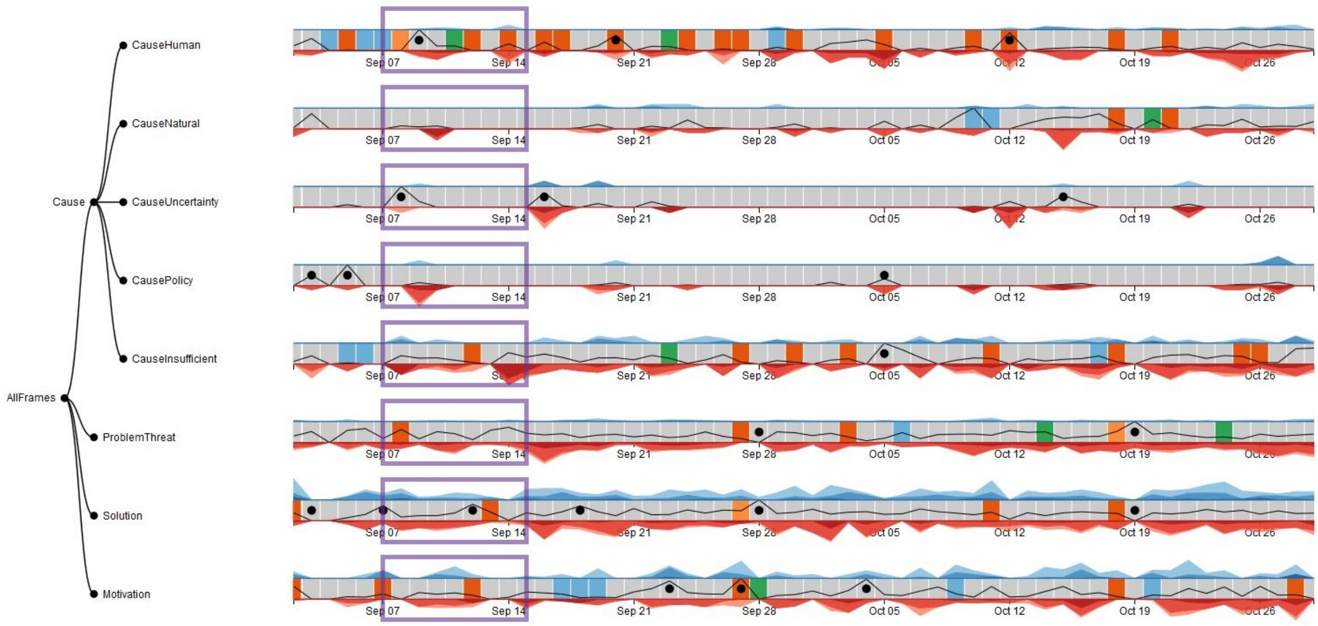


Fig. 4: Hierarchical timeline view showing intervention modeling results, Before-During-After analysis results and sentiment river for each expanded frame or frame category are shown. The frame structure is displayed as a dendrogram on the left. Clicking on the node can expand/collapse its children. The timeline associated with each leaf node is shown on the right.

Mathematically, when a time series model is affected by another input time series, a transfer function-noise model can be used to improve the model. The general form of this type of model is:

$$y_t = v(B)x_t + N_t \tag{1}$$

where y_t is the time series of interest, $v(B)$ is an autoregressive, integrated, moving average (ARIMA) model for the time series y_t, x_t is the input time series, and N_t is a noise process [31]. A specific case of transfer function-noise models is an intervention model, where the input time series is an indicator variable that specifies whether some event has taken place at time t. Such an event may have a temporary (or permanent) effect on the level or mean of the time series of interest.

An intervention model can model the effect of a known event on the time series. However, another common application of intervention models is to identify outliers in the time series. In this case, we do not know the exact time period in which the event (outlier) has taken place. The transfer-function model for this application then becomes:

$$y_t = v(B)\varepsilon_t + \omega I_T^{(t^*)},\ \text{where } I_t^{(t^*)} = \begin{cases} 1 \text{ if} & t = t^* \\ 0 \text{ if} & t \neq t^* \end{cases} \tag{2}$$

where ω is the change in the mean of the time series at time t^* and $I_t^{(t^*)}$ is an indicator function assuming that the effect of the outlier is temporary and only occurs at time period t^*. Other models can be used to model the case where an outlier may have a lasting impact on the mean of the time series. An iterative procedure is used to identify multiple outliers in the time series. In this scenario, multiple intervention models are fit, updating $I_t^{(t^*)}$ for $t^* = 1,\ldots,N$ for a time series with N time periods.

For the media data explored in this paper, intervention models were used to detect outliers, i.e. cues to events that may be of interest to the analyst, for each of the 25 frames over the time period of August 2 to December 31, 2014. For our intervention model, Figure 4 shows the trend of several frame categories. A black dot represents a statistically significant shift in frames between the week before and after this date. Users can then use the coordinated views to explore events that occurred at this time and begin forming hypotheses on the impact that events may have had on the media framing. Note that this is for event cueing and hypothesis generation. Events and frames of interest found require further investigation. Initial analysis of each time series (each frame) indicated that there was no significant autocorrelation present. Therefore, the intervention model can be simplified to:

$$y_t = \mu + \omega I_t^{(t^*)} + \varepsilon_t \tag{3}$$

where μ represents the overall mean of the time series and ε_t represents the error. Outliers at time $t^*, t^* = 1,\ldots,N$, can be identified by comparing the estimated value of ω, $\hat{\omega}$, to its standard error [31]. A significance level of $\alpha = 0.05$ was used to determine whether the value of the frame at time t^* was an outlier. The presence of an outlier cues the analyst to investigate what caused this change in the frame distribution. Although the intervention model is simplified because the frame time series were not autocorrelated, this approach is still valid for time series data that does have autocorrelation and Equation 2 would be used in such cases. Such models are sensitive to the time period under exploration. In this case, our analysts were exploring short term changes (1 week prior to the event, 1 week after the event). As such, the results of the intervention model tend to highlight peaks in the data; however, this is likely an artifact of the chosen window sizes. Future work will explore visual representations for exploring interventions under varying window sizes.

Before-During-After Analysis Since there was no autocorrelation in the data, a secondary model which requires an assumption of data independence, can be applied. The second intervention test defines a Before, During, and After period (where the during period can be seen as the intervention) and tests their location based on the data distributions. We let t denote the start time of the During period, and the time windows for the Before, During and After segments are represented by W_B, W_D, and W_A respectively. In this manner, the three time segments cover the following time periods: $Before: (t - W_B, t)$, $During: (t, t + W_D)$, and $After: (t + W_D, t + W_D + W_A)$. The data samples for the three segments are denoted as $D_B = \{x_1, x_2, \ldots, x_{n_B}\}$, $D_D = \{y_1, y_2, \ldots, y_{n_D}\}$, $D_A = \{z_1, z_2, \ldots, z_{n_A}\}$ and they may vary in length. Each data sample is the percentage of the frame in one document. Because there was no significant autocorrelation present in our underlying dataset, each sample is assumed to be independent and identically distributed ($i.i.d.$) where $D_i \sim N(\mu_i, \sigma_i^2)$. Therefore we form the problem to be tested as follows:

A_1: μ_D is not significantly different than μ_B
A_2: μ_D is not significantly different than μ_A
H_0: μ_D is not associated with an intervention $(A_1 \cap A_2)$
H_1: μ_D is associated with an intervention $(\bar{A_1} \cup \bar{A_2})$

We test H_0 by testing A_1 and A_2. To test A_1 and A_2, we applied a two-sample location test, Welch's *t-test* [40], on D_B, D_D and D_D, D_A individually with significance level α. In these two t-tests, the statement is the null hypothesis. Because of the multiple comparisons problem (in our case we have two tests one for D_B and D_D, and another for D_D and D_A), and based on the Bonferroni inequality $P(A_1 \cap A_2) \geq 1 - 2\alpha$, we applied Bonferroni correction and set the significance level for each test according to the following equation [32],

$$\alpha = \frac{\alpha_F}{\#test}, \tag{4}$$

where α is the significance level for each two-sized *t-test*, α_F is the family significance level for the multiple comparison for each During time period, and $\#test$ is the number of tests applied at each time period. In our case, $\#test$ equals 2 (the tests of A_1 and A_2). We set $\alpha_F = 0.05$, which guarantees that the overall significance level for the 2 hypothesis tests at each frame period is 0.05. To guarantee $\alpha_F = 0.05$, we set $\alpha = 0.025$ for each single test on the pair of consequent segments. Given the test result and the estimated μ, we can form 9 types of volume change patterns listed in Table 1. The 9 types are visualized in different color blocks on the time line for each frame, as shown in Figure 4 and Figure 5.

The color scheme also denotes the group of patterns as increasing, decreasing and oscillating. To change the interval length of each test time period, the user can change the size of the three windows for Before During and After using the spinners on the left. To better focus on a particular Before-During-After pattern, the user can click on the pattern legend to gray out options. In addition to knowing the intervention time point from the results of the intervention model, the Before-During-After analysis provides an adjustable window size and shows any significant changes.

The statistical tests' results are visualized in our timeline view for each frame and frame categories, shown in Figure 4 and Figure 5. The intervention modeling result is a set of binary indicators denoting the statistically significant intervention points. This result is represented as a black dot on our timeline view. The Before-During-After analysis's result is a set of patterns describing statistically significant changes in frame distribution over time. In both cases, the analyst can adjust the before, after and intervention (during) periods using the controls seen in Figure 1 (lower left). In our case study, the analysts were interested in a single day intervention with a 7 day news cycle.

Table 1: The pattern summary for Before-During-After analysis. Each pattern is associated with a unique hue as shown in the lower lefthand legend of Figure 1

pattern	sketch	description
$B = D = A$		no significant change
$B = D < A$		
$B < D = A$		increase
$B < D < A$		(blues)
$B = D > A$		
$B > D = A$		decrease
$B > D > A$		(greens)
$B > D < A$		oscillating
$B < D > A$		(oranges)

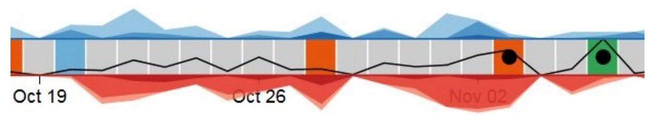


Fig. 5: Sentiment stacked area chart on bi-side of the time series view. The blue area represents positive sentiment and the red area river represents negative sentiment. The darker the area color is the more certain the label is for those sentences' sentiment class.

3.6 Frame Sentiment Visual Analytics

The underlying sentiment of the media and its relation to the framing can also provide insight. Sentiment analysis visualization has been successfully applied across a variety of domains, such as political election analysis [38], and merchandise reviews [33]. However, most classifiers are text context sensitive and need to be trained on a particular domain's data to boost performance. Furthermore, the limitation of sentiment classification accuracy is a problem in sentiment analysis and is subject to uncertainty [10]. In our visual analytics framework, we employ anl entropy-based sentiment river to reveal the uncertainty of sentiment over time using an ensemble voting scheme from multiple classifiers to determine the final sentiment label [26].

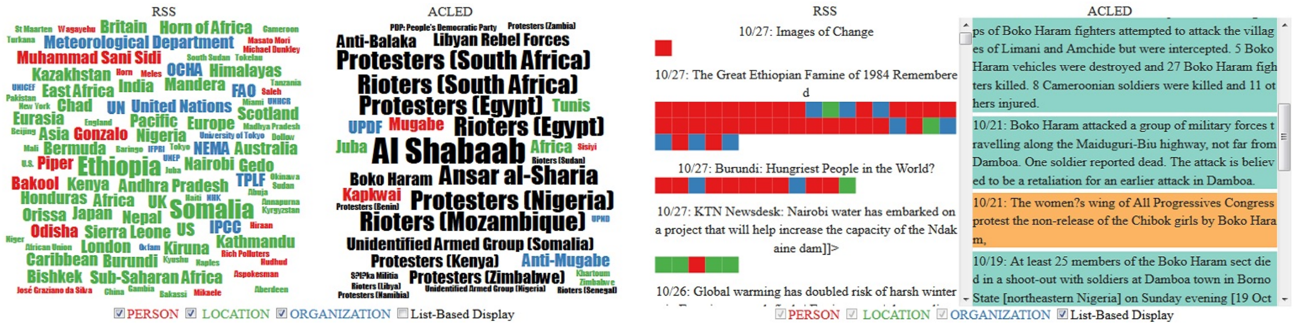
In our previous work [26], the uncertainty was visualized in each time period along the entropy sentiment river. However, the time information associated with RSS media data is not as precise as online social media data, such as Twitter. In general, the time parsed out from the RSS news is at the granularity of a day. In one day, there can be multiple articles collected relating to the target topic and each article also contains several frame coded sentences. Instead of exploring only the change of the certainty over the media stream, the volume of certain and uncertain sentiment labels is also explored. To enhance the understanding of the volume change for both certain and uncertain sentiment labels, a stacked area graph is used to represent each uncertain level with a stacked area and low uncertain area is stacked at bottom. Figure 5 shows this view, where the positive sentiment is colored in blue on top of the time series, while the negative sentiment is colored in red on bottom of the time series. The volume of relatively certain sentiment values are shown with a darker color and the uncertain volume with a lighter color. The height of the stacked area graph shows the average volume of sentences per document in each sentiment polarity over time as well as the portion of uncertainty. In this way, we can explore the positive and negative sentiment of the media documents in conjunction with their underlying frames.

3.7 Detail View

The detailed view, Figure 6, contains two modes, the entity wordle display and the list-based summary display. The data under analysis for this view changes along with the time period selection, the subset data selection for both media data and the secondary data, and the geospatial selection. When a user is only exploring the frame class 'Problem' which is colored in red, only the RSS articles containing at least one sentence being framed as 'Problem' will be displayed in the detail view. For a geospatial selection, e.g. the user selects a country to explore, the data displayed in the detail view updates to filter for only the articles and ACLED events related to this country.

In the entity wordle view, the most frequently named entities extracted from the two datasets are displayed in two wordles. Based on the entity's class, which is either person, location, or organization, the word is colored in red, green, or blue respectively. The actors in the ACLED dataset, being entities as well, are colored in black. The size of those entities displayed here is also proportional to its frequency.

In the list-based summary view, the RSS news article is summarized by showing the title and a list of colored squares, where each square represents each framed sentence colored by its frame class's color. The ACLED data, being the secondary data here, displays its notes for each event in the selected time period. The background color of each note matches the color for its event type. To analyze events containing



(a) Detail view in entity wordle display (b) Detail view in list-based summary display

Fig. 6: The detail view showing the most frequently named entities in a wordle display and document summary information in a list-based display. Here we show data from Oct.11 to Oct 27. View (a) is the entity wordle display in which user can choose three classes of entities (person, location, organization) to show. Black text in the ACLED wordle indicates an actor in the events. View (b) is the list-based summary display in which the title of media articles and the summary information of the secondary dataset are listed in order of time. The frame information of each article is summarized into colored squares (the color of the square matches the frame class) in the sentence order from the article. In this example, the ACLED event notes are filtered by clicking on the entity text 'Boko Haram'.

a particular entity of interest, a user can click on a particular entity shown in the wordle display and information containing that entity will show up in the summary display. Users may also filter by location by selecting a country in the geographical view.

4 Case Study: Climate Change Framing and Armed Conflict in Africa

In this section, we demonstrate our work by applying the methods described so far to the RSS news dataset collected on Climate Change from African countries and the ACLED data set, which focuses on armed conflicts and political violence in Africa. Collaborators were interested in linking these two data sets based on previous articles and reports that discussed the impacts of climate change on armed conflicts [34]. Their goal was to explore the framing of news stories related to climate change and see what, if any, armed conflict events may be linked to that discourse. In this manner, social scientists can begin to develop models and theories about how framing can help drive political change, or conversely, how armed conflict is driving discourse.

4.1 Exploring Problem Frames in Africa

The analyst begins with an overview of the system and explores the distribution of frames over the entire time period of data collection. The main points of interest are the spatial and temporal distributions of frames, Figure 7. First, the analyst explores the spatial distribution of frames, looking at the weighted majority choropleth map. The analyst notes that most regions are discussing climate change either in terms of problems (red) or solutions (green). Only a few countries, such as the Republic of Guinea-Bissau and the Republic of Côte d'Ivoire, have a majority distribution related to causes of climate change, and Congo has more motivation frames. The analyst drills down into the data by selecting a country and quickly learns that only one document has geographic information related to these countries. Thus the analyst determines that these outliers are of little interest.

Given that the discourse seems to focus heavily on both problems and solutions, the analyst decides to explore the temporal view with a focus on problems. The analyst searches the top level problem hierarchy looking for significant events found in both the intervention model and Before-During-After model. The analyst finds a time period in late October (highlighted as circle **a** in Figure 7) with several points of interest, and highlights this time period for inspection. The analyst then expands the tree and explores the leaf node problem frames, Figure 7(bottom). The analyst notes that there are significant interventions in many categories, but very few frames regarding security threats and water problems in this time period. The analyst further comments on the lack of water framing in the documents noticing that climate change is often associated with extreme weather, including drought, yet there is little discussion in Africa about problems related to water. The analyst does notice that there are many documents discussing problems with food.

The analyst decides to first focus on the food problem frame and the events leading to this change in the frame distribution. The analyst narrows the time period to October 11th to October 28th, and then filters for RSS news articles containing frame category 'ProblemThreat' and ACLED events Riots and Battles. The analyst wants to explore what geographic regions are seeing large amounts of armed conflict during this time period. The analyst selects the most prevalent ACLED events (Riots-yellow and Battles-red) using the donut control. The analyst notes that the largest amount of conflicts are occurring in Nigeria, Sudan and Somalia. Given the importance of the Niger River Basin in the area, the analyst chooses to explore events in Nigeria that may be driving the discourse on climate change. The analyst notes that it is interesting that there is a clear separation between the riots (in the south) and the battles (in the north). The analyst selects Nigeria to filter the detailed view to only RSS documents and ACLED events that are geocoded to Nigeria.

Looking at the RSS articles' titles, the analyst finds many reports talking about the problem of food security and famine in Africa. While exploring the ACLED events in the same time period, the analyst locates several riots and battles discussing the impact of Boko Haram on farmers, where militarists are killing farmers and forcing them to flee their homes, exacerbating the food problem. Example articles and events are shown in Figure 8(Left). What is interesting to the analyst is that articles are already discussing the famine problems that Africa will face due to climate change. If this is further exacerbated by wars, the problem cycle may become more prevalent resulting in displacement, migration, and potential social unrest. From a social science perspective, our analysts are interested in how to model such phenomena. By cueing them to such events, they are able to begin looking at how ongoing events could be modeled to predict future problems.

After discussing the events surrounding the food frame cue, the analyst then decides to also explore the ProbThreatHealth frame (problems associated with health) next. The analyst is interested in the two significant events that occurred between October 11th and October 28th. Again, the analyst begins exploring related ACLED events during this time period, and quickly finds several riot/protest events related to the mistreatment of healthcare workers in the region. The analyst again noted their interest in these articles and the fact that the event cueing was able to narrow down their search to potentially relevant information. While there are some obvious links between food security, armed conflicts and riots (for example, Boko Haram displacing farmers), subtle social issues involved with riots may be harder to spot. Furthermore, given that such riots are taking place at this time and there is a shift in frames, the analyst hypothesized that this could represent a shift in the discourse in the hopes to alleviate concerns

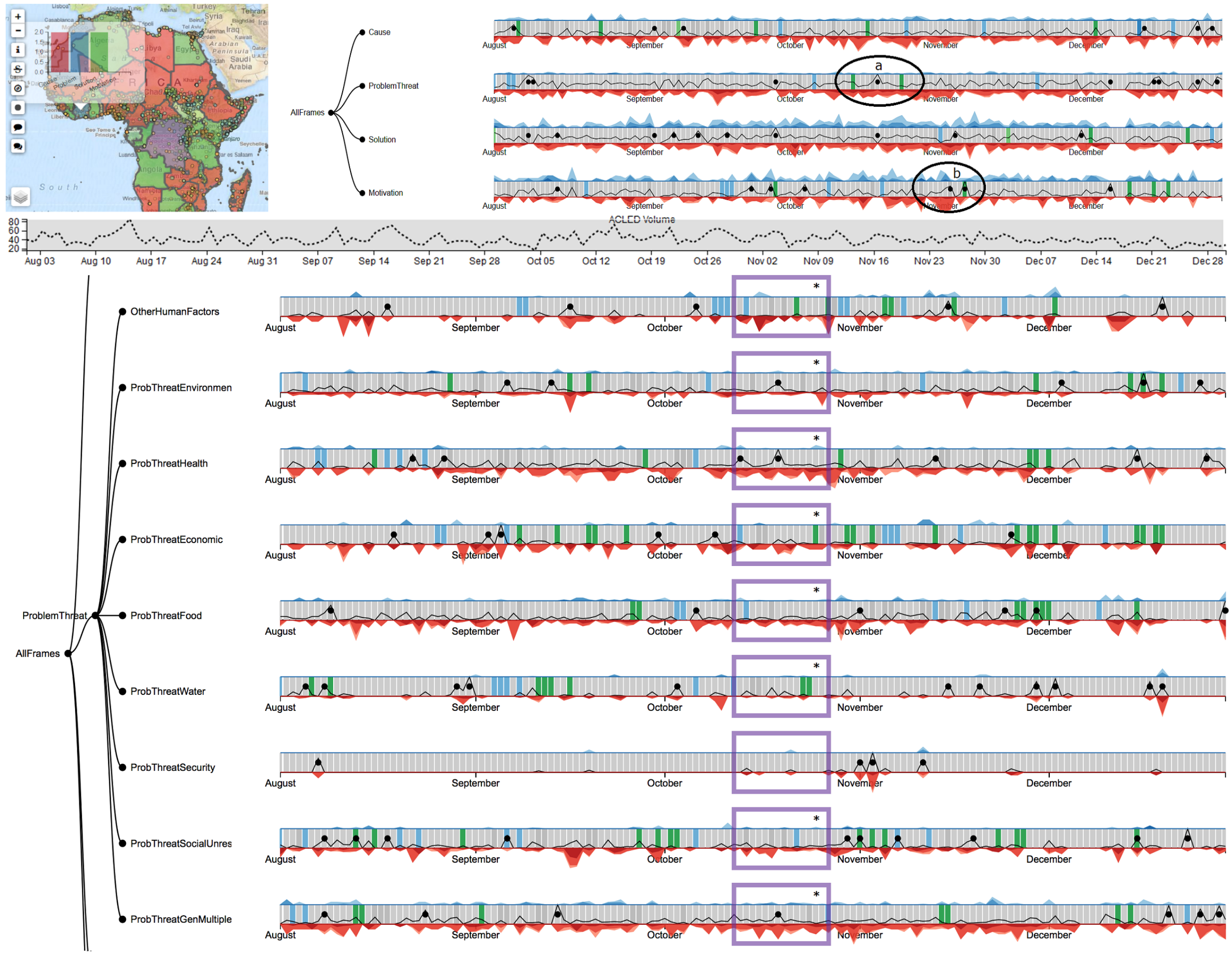


Fig. 7: Exploring the whole time period on the RSS news dataset spatially and temporally. The spatial map shows a weighted choropleth map with all frame class equally weighted. The add-on histogram shows the frame volume and distribution of the Republic of Côte d'Ivoire. The top timeline view shows the level of four frame classes and two black circles highlight the time period of interest in ProblemThreat and Motivation. The bottom timeline view shows the expanded timelines in the ProblemThreat frame class and the time period of interest is highlighted.

from the general population. While no definitive conclusions could be made at this time, this example further illustrates how our framework can enhance the hypothesis generation process. By specifically cueing an analyst to a time of interest, we can dramatically cut their exploratory analysis time. For example, there are over 40 ACLED events per day, each with an associated set of documents. Uncued analysis of such work would be an extremely laborious process.

4.2 Exploring Motivation Frames in Africa

The analyst concentrates on examining press coverage between November 1st and November 14th, and identifies events accounting for notable intervention points on November 6th based on the Before-During-After model. Results indicate an increasing trend in the media discourse on calling for policy actions on November 2nd with a negative tone. The statistically significant interventions and the burst of the sentiment can be found in the Figure 7(highlighted in circle **b**). The changing pattern is predominantly associated with the launch of an updated synthesis report by the UN's Intergovernmental Panel on Climate Change (IPCC) on November 2nd. Several articles reporting IPCC can be easily found and shown in Figure 9. As the most comprehensive assessment that attracts worldwide attention, the new IPCC report summarizes alarming evidence detailing severe impacts of climate change. Adverse impacts range from increased risks of extreme weather events, food shortages, and violent conflicts. The alarming messages, circulated by several media outlets, were framed in mostly negative words (e.g. serious impacts, severe impact, dangerous, catastrophic). In addition, analysts find prevalent explicit statements calling for international governments to take actions now. The following sentences describe examples of motivational framing.

- "Massive cuts to greenhouse gas emissions are needed in the coming decades to curb temperature rises to no more than 2C, the level at which it is thought dangerous impact of climate change will be felt."
- "Leaders must act."
- "There is cause for hope if governments take action."
- "A binding meaning and enforceable framework is needed to limit the consequences of global warming."'
- "The world's largest polluters, the United States and China, should take the lead in reducing emissions."

Conversely, there are noticeable spikes of positive sentiment values between November 9th and November 12th. The pattern is largely associated with favorable coverage of U.S. and China announcing a historical climate change agreement on November 11 when President Obama visited Beijing for the Asia Pacific Economic Cooperation (APEC) summit. Together, motivation frames in West Africa reflect a focus of relying on international actors to drive policy negotiation.

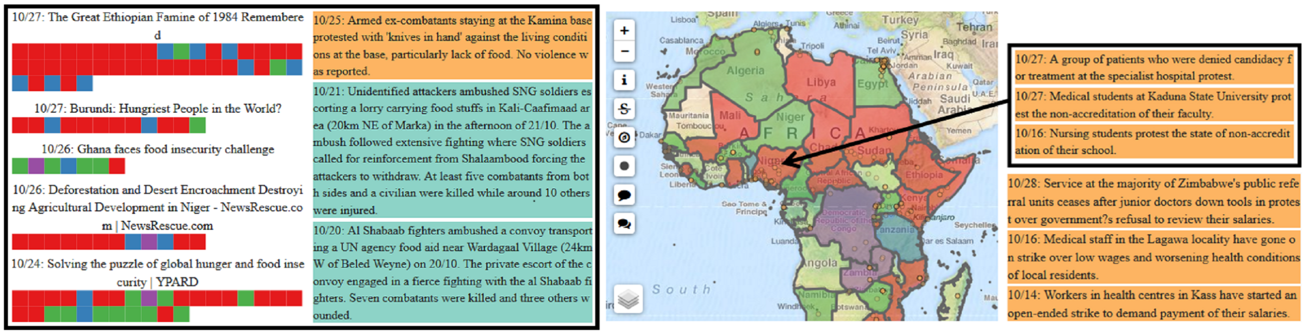


Fig. 8: The geographical and detail view for exploring RSS news and ACLED data. This figure shows the analyzing time from October 11th to October 28th. The geographical view color each country by the majority frame class and displays riots (orange dots) to represent the ACLED events. The detail view lists the Riots events related to health problem within and outside Nigeria. The left side detail view shows examples of RSS articles discussing food problem in Africa and the ACLED events are riots and battles expressing problem of food supply.

Results of analysis on motivational frames should be viewed in light of limitations. In the 1,245 relevant articles collected from West African news media and twitter links, there is little evidence of motivation framing, as less than 10% of a news story contained statements calling for definitive courses of actions, That is, motivational frames are very uncommon compared to other three frame classes (cause, problem/threat, and solution). When a set of news stories highlighted explicit calls for actions to solve climate change issues within the same time period, it is highly possible that the consistent pattern in press coverage was statistically significantly different than before and after in the time series analysis. Despite the low presence of motivational statements in the current dataset, the visualization tool allows researchers, analysts, and policy makers to explore the potential underlying mechanisms linking adverse impacts of climate change and increased risk of political conflicts.

4.3 Analyst Feedback

Our case study involved two analysts from the Department of Communication at Arizona State University. Feedback on the system was positive with analysts indicating that the event cueing features were extremely useful in providing a starting point for searching linked data. Case Study 1 was done as a paired analysis demonstrating the tool with the computer scientists manipulating the controls and discussing how the system worked. Case Study 2 was done at the communication lab with no assistance from the computer science group (the tool is web-deployed).

Overall feedback was positive with the analysts stating that they were "fascinated by the visualization tool's ability to map out temporal and spatial components of media discourse". In addition, the analysts also mentioned that this tool can help to tackle co-occurrence patterns of conflicting events, limiting the possibility of bridging distinct lines of scholarship together–media research, climate change and conflicts. However, there were suggestions for future work and improvement. Specifically, the analysts were interested in the difference between the change models and their disagreements. For example, in Figure 1, there is an intervention marker (black dot) near October 5th for motivation, but no colored squares from the before-during-after analysis. The relationship between these two models required more explanation and future work will explore creating a single ensemble metric. Along with the intervention model, the analysts also requested the ability to reconfigure layouts for improved storytelling. They indicated that they would be able to better explore relationships with a series of small multiples and better alignment between the temporal components of the unrest data and the framing data.

5 Conclusions and Future Work

In this paper, we have demonstrated a framework for event cueing that enables the exploration of evolving media discourse. Our framework focuses on both the spatial and temporal distribution of frames, and allows experts to quickly explore spatial trends in the underlying discourse. By linking multisource data for exploration, our framework

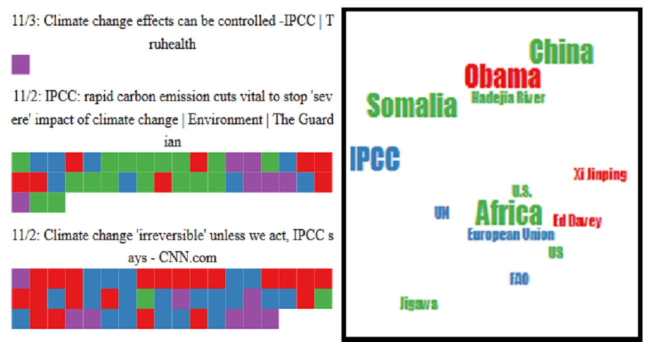


Fig. 9: Example RSS articles and the entity wordle for the time period of Oct. 28th to Nov. 11th exploring motivation frame. The left side article summaries show examples of news report relating to the IPCC and the right side wordle emphasizes the most frequent entities appearing in those articles, such as IPCC, Obama, and China.

enables users to explore more complex hypotheses that can enable analysts to link potential cues between disparately collected sources. While several visual analytics methods [11, 12, 13] have explored frames in the context of comparing corpora of text and topical terms within these text, our framework enables sentiment analysis and intervention modeling which can provide different insights than previous work.

Our framework was evaluated through collaboration with domain experts from the School of Communication and findings from their exploration have prompted new questions and directions to explore. While our examples focused on climate change and conflicts in Africa, the tools developed are applicable for a variety of media sources. Furthermore, it is important to note that our intervention strategy can be applied to any temporal variable, and, by utilizing multiple models, we are able to strengthen the analysts' confidence in the findings. This was particularly evident in the exploration process. Anomaly detection methods, intervention models and others often have a large false positive rate. By using an ensemble of models, one can begin defining uncertainty. Future work will focus on a combination of anomaly models and intervention models as well as a weighted output for defining uncertainty in the detection, similar to our sentiment modeling approach. We also plan to explore a combination of sentiment analysis, frames and clustering for defining geo-political regions that share common framing strategies. We believe that such methods can further enable multisource data exploration and provide new cues to analysts who are developing hypotheses and exploring the evolution of topics, events and discourse both locally and globally.

Acknowledgement

Some of the material presented here was sponsored by Department of Defense and is approved for public release, case number 15-365 and upon work supported by the NSF under Grant No. 1350573.

REFERENCES

[1] Armed conflict location & event data project. `http://http://www.acleddata.com/`. Accessed: 2015-03-28.

[2] Data science toolkit. `http://www.datasciencetoolkit.org`. Accessed: 2015-03-18.

[3] W. Aigner, S. Miksch, H. Schumann, and C. Tominski. *Visualization of time-oriented data*. Springer Science & Business Media, 2011.

[4] R. D. Benford and D. A. Snow. Framing processes and social movements: An overview and assessment. *Annual Review of Sociology*, pages 611–639, 2000.

[5] M. Bogl, W. Aigner, P. Filzmoser, T. Lammarsch, S. Miksch, and A. Rind. Visual analytics for model selection in time series analysis. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2237–2246, 2013.

[6] H. Bosch, D. Thom, M. Worner, S. Koch, E. Puttmann, D. Jackle, and T. Ertl. Scatterblogs: Geo-spatial document analysis. In *Proceedings of IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 309–310. IEEE, 2011.

[7] J. Chae, D. Thom, H. Bosch, Y. Jang, R. Maciejewski, D. S. Ebert, and T. Ertl. Spatiotemporal social media analytics for abnormal event detection and examination using seasonal-trend decomposition. In *Proceedings of IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 143–152. IEEE, 2012.

[8] V. Chandola, A. Banerjee, and V. Kumar. Anomaly detection: A survey. *ACM Computing Surveys (CSUR)*, 41(3):15, 2009.

[9] D. Chong and J. N. Druckman. Framing theory. *Annual Review of Political Science*, 10:103–126, 2007.

[10] N. F. da Silva, E. R. Hruschka, and E. R. Hruschka. Tweet sentiment analysis with classifier ensembles. *Decision Support Systems*, 66:170–179, 2014.

[11] 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*, 2015.

[12] N. Diakopoulos, A. Zhang, D. Elgesem, and A. Salway. Identifying and analyzing moral evaluation frames in climate change blog discourse. In *Proceedings of International Conference on Weblogs and Social Media (ICWSM)*, 2014.

[13] N. Diakopoulos, A. X. Zhang, and A. Salway. Visual analytics of media frames in online news and blogs. In *Proceedings of IEEE InfoVis Workshop on Text Visualization*, 2013.

[14] 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 *IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 93–102. IEEE, 2012.

[15] W. Dou, L. Yu, X. Wang, Z. Ma, and W. Ribarsky. Hierarchicaltopics: Visually exploring large text collections using topic hierarchies. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2002–2011, Dec 2013.

[16] R. M. Entman. Framing: Toward clarification of a fractured paradigm. *Journal of Communication*, 43(4):51–58, 1993.

[17] P. Federico, S. Hoffmann, A. Rind, W. Aigner, and S. Miksch. Qualizon graphs: Space-efficient time-series visualization with qualitative abstractions. In *Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces*, pages 273–280. ACM, 2014.

[18] D. Fisher, A. Hoff, G. Robertson, and M. Hurst. Narratives: A visualization to track narrative events as they develop. In *IEEE Symposium on Visual Analytics Science and Technology*, pages 115–122. IEEE, 2008.

[19] T. Gao, J. R. Hullman, E. Adar, B. Hecht, and N. Diakopoulos. Newsviews: An automated pipeline for creating custom geovisualizations for news. In *Proceedings of the 32nd annual ACM conference on Human Factors in Computing Systems*, pages 3005–3014. ACM, 2014.

[20] D. Gotz and H. Stavropoulos. Decisionflow: Visual analytics for high-dimensional temporal event sequence data. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1783–1792, 2014.

[21] J. Hullman, N. Diakopoulos, and E. Adar. Contextifier: Automatic generation of annotated stock visualizations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 2707–2716, New York, NY, USA, 2013. ACM.

[22] S. Ko, S. Afzal, S. Walton, Y. Yang, J. Chae, A. Malik, Y. Jang, M. Chen, and D. Ebert. Analyzing high-dimensional multivariate network links with integrated anomaly detection, highlighting and exploration. *Proceedings of IEEE Conference on Visual Analytics Science and Technology*, pages 83–92, 2014.

[23] M. Krstajic, E. Bertini, and D. A. Keim. Cloudlines: Compact display of event episodes in multiple time-series. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2432–2439, 2011.

[24] M. Krstajić, M. Najm-Araghi, F. Mansmann, and D. A. Keim. Story tracker: Incremental visual text analytics of news story development. *Information Visualization*, 12(3-4):308–323, 2013.

[25] S. Liu, X. Wang, J. Chen, J. Zhu, and B. Guo. Topicpanorama: A full picture of relevant topics. In *Proceedings of IEEE Conference on Visual Analytics Science and Technology*, pages 183–192. IEEE, 2014.

[26] Y. Lu, X. Hu, F. Wang, S. Kumar, H. Liu, and R. Maciejewski. Visualizing social media sentiment in disaster scenarios. In *Proceedings of the 24th international conference on World Wide Web companion*. International World Wide Web Conferences Steering Committee, 2015.

[27] 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, 2012.

[28] A. Malik, R. Maciejewski, N. Elmqvist, Y. Jang, D. S. Ebert, and W. Huang. A correlative analysis process in a visual analytics environment. In *Proceedings of IEEE Conference on Visual Analytics Science and Technology*, pages 33–42. IEEE, 2012.

[29] C. D. Manning, M. Surdeanu, J. Bauer, J. Finkel, S. J. 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*, pages 55–60, 2014.

[30] M. Monroe, K. Wongsuphasawat, C. Plaisant, B. Shneiderman, J. Millstein, and S. Gold. Exploring point and interval event patterns: Display methods and interactive visual query. *University of Maryland Technical Report*, 2012.

[31] D. C. Montgomery, C. L. Jennings, and M. Kulahci. *Introduction to Time Series Analysis and Forecasting*. Hoboken, NJ: John Wiley & Sons, 2008.

[32] J. Neter, M. H. Kutner, C. J. Nachtsheim, and W. Wasserman. *Applied linear statistical models, 5th edition*, volume 4. Irwin Chicago, 1996.

[33] D. Oelke, M. Hao, C. Rohrdantz, D. Keim, U. Dayal, L. Haug, and H. Janetzko. Visual opinion analysis of customer feedback data. In *IEEE Symposium on Visual Analytics Science and Technology*, pages 187–194, 2009.

[34] J. OLoughlin, A. M. Linke, and F. D. Witmer. Effects of temperature and precipitation variability on the risk of violence in sub-saharan africa, 1980–2012. *Proceedings of the National Academy of Sciences*, 111(47):16712–16717, 2014.

[35] D. A. Scheufele. Framing as a theory of media effects. *Journal of Communication*, 49(1):103–122, 1999.

[36] D. Thom, H. Bosch, S. Koch, M. Worner, and T. Ertl. Spatiotemporal anomaly detection through visual analysis of geolocated twitter messages. In *Pacific Visualization Symposium*, pages 41–48. IEEE, 2012.

[37] J. J. Van Wijk and E. R. Van Selow. Cluster and calendar based visualization of time series data. In *IEEE Symposium on Information Visualization*, pages 4–9. IEEE, 1999.

[38] F. Wanner, C. Rohrdantz, F. Mansmann, D. Oelke, and D. A. Keim. Visual sentiment analysis of RSS news feeds featuring the US presidential election in 2008. In *Workshop on Visual Interfaces to the Social and the Semantic Web*, 2009.

[39] F. Wanner, A. Stoffel, D. Jäckle, B. Kwon, A. Weiler, D. Keim, K. E. Isaacs, A. Giménez, I. Jusufi, T. Gamblin, et al. State-of-the-art report of visual analysis for event detection in text data streams. In *Computer Graphics Forum*, volume 33, 2014.

[40] B. L. Welch. The generalization of 'student's' problem when several different population variances are involved. *Biometrika*, pages 28–35, 1947.

[41] K. Wongsuphasawat and D. Gotz. Exploring flow, factors, and outcomes of temporal event sequences with the outflow visualization. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2659–2668, 2012.

[42] K. Wongsuphasawat, J. A. Guerra Gómez, C. Plaisant, T. D. Wang, M. Taieb-Maimon, and B. Shneiderman. Lifeflow: Visualizing an overview of event sequences. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1747–1756. ACM, 2011.

[43] C. Xie, W. Chen, X. Huang, Y. Hu, S. Barlowe, and J. Yang. VAET: A visual analytics approach for e-transactions time-series. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1743–1752, 2014.