Visual Event Cueing in Linked Spatiotemporal Data

by

Michael Steptoe

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved October 2017 by the
Graduate Supervisory Committee:

Ross Maciejewski, Chair
Hasan Davulcu
Steven Corman

ARIZONA STATE UNIVERSITY

December 2017

ABSTRACT

The media disperses a large amount of information daily pertaining to political events social movements, and societal conflicts. Media pertaining to these topics, no matter the format of publication used, are framed a particular way. Media framing can be defined as a form of communication that presents information to its audience in a way that encourages a particular point of view [12]. Framing is used not for just guiding audiences to desired beliefs, but also to fuel societal change or legitimize/delegitimize social movements. Understanding media framing is necessary to analyze the stability or change in social climate [6]. For this reason, tools that can help to clarify when changes in social discourse occur and identify their causes are of great use. This thesis presents a visual analytics framework that allows for the exploration and visualization of changes that occur in social climate with respect to space and time. Focusing on the links between data from the Armed Conflict Location and Event Data Project (ACLED)[2] and a streaming RSS news data set, users can be cued into interesting events enabling them to form and explore hypothesis. This visual analytics framework also focuses on improving intervention detection, allowing users to hypothesize about correlations between events and happiness levels, and supports collaborative analysis. The results of Granger causality testing are reported to the user to emphasize if there is a causality relationship between ACLED events and RSS articles. A geographical view is used to map the sentiment and well-being of a region's population allowing users to form hypotheses about the impact events have on population happiness. To support collaborative analysis, a view is provided for users to comment on findings and share them with others and support for coordinated exploration is also provided.

DEDICATION

To my family and friends for all of their encouragement

ACKNOWLEDGMENTS

Foremost, I would like to thank my advisor and mentor Dr. Ross Maciejewski for providing me the opportunity to work in his lab. This thesis would not have been possible without his guidance and continuous support. I would also like to give thanks to Dr. Hasan Davulcu and Dr. Steven Corman for being a part of my graduate committee and their support throughout my research. This work was partially supported by the National Science Foundation, Grant No. 1350573 and the U.S. Department of Homeland Security's Vaccine Center under Award No. 2009-ST-061-CI0001.

TABLE OF CONTENTS

LIST OF FIGURES

Figure Page

Chapter 1

INTRODUCTION

Framing in the media has a direct impact on the stability of social movements guiding them towards success or failure [6]. Frames evolve as the perceived reality of the speaker and the point of view they wish to convey changes. Event Cueing aims to help users understand how frames are applied, when changes in framing occur, and to identify the events that may have aided the change. These events can help to explain the change in happiness across different regions and add depth to the media's framing. Understanding the causes of conflicts, their impacts, and the drivers behind them is a major project in the social sciences [35]. In this work, we present a framework designed to helps users work towards identifying and understanding these issues.

This thesis presents the Event Cueing framework. Utilizing intervention models for time series analysis, we can cue users to time periods of interest. From these time periods, users can explore frame evolution through the secondary linking of ongoing events. Using predictive causality, the framework can help analysts determine if there is a causal relationship between the ACLED events and RSS articles. Our framework enables user's to explore the happiness levels of countries in Africa using their well-being and sentiment to understand when and how events shift population happiness. Users are able to make comments to tag their findings and label their evidence. Users can also explore, cooperatively sharing comments as they go through the system or inspect another user's findings on their own to help support or disprove their hypothesis.

This work is related to previous works in statistical analysis, spatiotemporal analysis, and collaborative analysis. Intervention modeling is applied to detect outliers in the event data similar to work by Bogl et al. [8], which uses autoregressive integrated moving average (ARIMA) models. Through the use of this intervention modeling, Event Cueing can cue users to events in time series datasets. Using the time periods of interest from the intervention models, users can explore the events that occurred during that time and hypothesize about causes of frame changes. By testing for Granger casuality, our framework can determine if the ACLED dataset has a predictive causal relationship with the RSS dataset over the period of interest. If the relationship exists, then the dataset may contain useful information for predicting future changes in frames. This analysis can be used as an indicator for users and help them determine trends and drivers of interest. Leveraging this automated analysis, users can quickly detect and explore suspected impacts and drivers of social change without performing manual preliminary investigations to identify them. For each frame category, sentiment analysis is applied to explore the change of certainty and the volume of certain and uncertain sentiment labels.

Extending the application of sentiment analysis, this work uses the collection of Tweets within a defined time period to determine the well-being for a given region. Past and current studies on well-being use polling and surveys to determine an index of geographic happiness. Our focus is specifically on using streaming media data, such as RSS news articles to measure the happiness of a given region. Social media has been successfully used to analyze temporal patterns of happiness and its geographical patterns within twitter data. Dodds et al. [23] propose a method for creating a tunable metric for evaluating the happiness of texts using a frequency-based word list. Mitchell et al. [52] extended upon this work by applying the word list to a geo-tagged

Twitter data set to estimate the happiness levels of states and cities. Using sentiment analysis, users are able to following the emotional discourse related to different frame categories and understand how these emotions are spatially distributed. In their hypothesis process, they can link events to changes in emotion across the geographical view and reason out why they may or may not expect a particular outcome.

Event Cueing allows users to quickly find events of interest and gather evidence to make hypothesis. In order to enhance the hypothesis generation process, this work extends collaborative analysis by incorporating shared views and visual annotation allowing for rapid result synthesis. Mahyar et al. [48] developed CLIP, which uses linked common work (LCW) to support collaboration and improve analytic outcomes. Work by Willet et al. [70] utilizes comments, tags, and links to help analysts classify evidence and establish common ground. Our work differs by providing live coordination between users during exploration, this means that users can work within the same space at the same time without having to load new configurations as snapshots are made. This accelerates user workflow processes, and helps users to focus on a particular task. At any point, users can return to explore findings of interest during their exploration and extend or develop new hypothesis based on these results.

This thesis addresses the challenges of identifying framing in media, exploring how media is framed over space and time, determining the drivers of change in media framing, exploring and supporting complex hypotheses, effectively combining multiple data sources, and collaborative visual analysis. While there has been previous work in the development of tools for frame analysis [19, 20, 21], none of these support entity extraction, sentiment analysis, linking to multi-source data, causality testing, or collaborative analysis. These additional analysis tools are necessary to better support the hypothesis generation process and to expedite the identification of interesting periods, entities, and drivers. This thesis focuses on both the spatial and temporal distribution of frames, allowing users to quickly explore spatial trends in the underlying discourse. Not unlike works such as Narratives[28] and EventRiver[46], which allow users to explore topics and keywords and associate them with other ongoing stories and events, this work moves to multi-source data utilizing model techniques described above. By building upon previous work [8, 48, 70] and through collaboration with domain experts, users can be cued in to potential links across collected sources and work collaboratively to strengthen their hypotheses.

The remainder of this thesis is organized in the following manner. Chapter 2 summarizes other works in visual analytics for media analysis, space time analysis, data linkage, and collaborative visual analysis. Chapter 3 gives an overview of the Event Cueing system, data collection and wrangling performed, and the visual analytic techniques utilized. Chapter 4 covers two case studies exploring media framing in Africa and provides feedback from analysts. Finally, in Chapter 5, the formal conclusion and future work are provided.

Chapter 2

RELATED WORK

Event Cueing uses a variety of text analysis tools and techniques to explore framing, including media analysis, geographic visualization, data linkage, and collaboration. In this section, I review past and current research in the following areas of visual analytics: media analysis; space time analysis; data linkage, and; collaborative visual analysis.

2.1 Visual Analytics for Media Analysis

Media data has been used to improve sensemaking [26], increase situational awareness [9], identify and follow political discourse [20]. Previous work on media data has focused on topic extraction, topic evolution, opinion diffusion, and coopetition analysis [14, 15, 61, 74].

Cui et al.[15] developed TextFlow, for the analysis of various evolution patterns that appear from multiple topics. TextFlow integrates visualization with topic mining techniques allowing users to refine their analysis and gain insights into topic evolution patterns. By utilizing an iterative and progressive visual analysis process they enable smooth communication between visualization and topic mining through interaction. Extending on previous work, Cui et al.[14] presented RoseRiver, which focuses on exploring and analyzing evolutionary hierarchical topics to guide users to insights about the topics leveraging an incremental evolutionary tree cut algorithm. RoseRiver solves the problems of flat topic assumption and the comprehension of hierarchical topic evolution by extracting topic alignments at the same level across trees and allowing users to progressively explore and analyze complex evolution patterns. Their visual analysis system connects large text corpora with people by presenting topics of interest with their evolution over time using a novel approach. Event Cueing utilizes an iterate visual analysis process similar to previous works providing the user meaningful interactions and cues from our text analysis techniques. Our system goes a step further and allows users to explore the evolution and hierarchy of frames both temporally and spatially.

EvoRiver [61] was designed to help users determine cooperation and competition relationships between topics using Twitter data sets by modeling complex interactions among topics. Through text analysis and modeling, EvoRiver is able to depict how different groups of influential users affect coopetition and allows users to explore and detect evolving patterns. Other applications of modeling and text analysis can be see in work such as OpinionFlow [74], which approximates opinion propagation among Twitter users at various levels using an opinion diffusion model. OpinionFlow, allows users to examine and compare difussion patterns across topics and drill down into the details to examine specific patterns. Our system uses text analysis techniques similar to previous works to perform various functions such as sentiment extraction, entity extraction, and frame extraction. Using the features extracted from these techniques we are able to model frame changes to enable analysts to detect and examine patterns.

In addition to supporting topic extraction and evolution, sentiment analysis of large text corpora is another growing research area. Examples of this research include the work of Gregory et al. [32], in their paper they implement interactive exploration of emotion in a large document collection. Their system enables users to compare the range of affect in documents within the collection and to relate affect to other dimensions within the documents support the analytical process. Wanner et al. [66] developed a visual analytics tool for semi-automatic sentiment analysis of large news feeds allowing analysts to efficiently gather meaningful information without reading all of the articles. Combining their sentiment analysis with news similarity filtering they allow the expert to focus on finding trends and putting events into context. Diakopoulos et al. [17] demonstrated visuals and metrics to help analysts better understand temporal patterns of sentiment from Twitter and later develops Vox Civitas [18], which uses sentiment patterns and content structure to extract prominence from social media outlets. Vox Civitas was designed to help journalists and media professionals extract news value from social media content around broadcast events. Work by Woldemariam [71] introduced the integration of a sentiment analysis pipeline into a cross-media analysis framework to extract semantics from the media in a cross-media context. In Woldemariam's work, there are comparisons between sentiment analysis methods and their appropriateness for detecting sentiment from explicit sources. In our system, sentiment analysis is used similarly to previous works to detect framing from media sources. Contributing these past works, we extend the use of sentiment analysis to prediction modeling and data filtering both spatially and textually.

Previous work on media frame visual analytics [19, 20, 21, 56], focused on topic analysis and comparison while our system focuses on frame analysis with secondary data linking. LingoScope [19] supports frame reflection and identification with relation to news producers and news consumers. Their work focuses on addressing the issue of comparison between sources, and comparison over time for a single or multiple sources. Diakopoulos et al. [21] also explored how numerous topics are framed in moral light and compared how they are discussed between acceptors and skeptics. Users could observe and filter the contextual terms that convey framing across large volumes of text and examine their details. Olteanu et al. [56] introduced a methodology for discovering triggers, actions, and news values through the analysis of a series of events. Their comparative analysis covers a span of 17 months demonstrating their method that combines automatic and manual annotations. Most similar to our work in frame analysis is Compare Clouds by Diakopoulos et al. [20]. In their work, they focus on the framing of a single topic and use this analysis to compare media discussion between mainstream media and blogs. Their system maps word prevalence and context information between the two sources of media information. In order to explore the evolution of frames, our system uses a hierarchical set to categorize and organize documents. By linking ACLED events we can help to cue users into points of interest relating to the changes in frames and guide their exploration through the topic discussion.

2.2 Visual Analytics for Space Time Analysis

Visualizing spatiotemporal data collections over their temporal or spatial counterparts enhances the analysis process improving the ability to discover patterns and recognize relationships across both domains. Past work in temporal event sequence analysis is similar to techniques we demonstrate in this work. Itoh et al. [37] proposed a system that extracts events from blog data and categorizes the events according to thematic roles, enabling users to explore temporal changes in events related to a topic. Their framework enables users to find events about a topic of interest that appears within a specified timing providing details about the events. Decision Flow [30] is a visual analytics tool that enables quick and accurate completion of sequence analysis tasks for temporal event sequence data. Decision Flow utilizes interactive multi-view visualizations and ad hoc statical analytics to support the analysis of high-dimensional temporal event sequence data. Wongsuphasawat et al. developed the Outflow visualization [72], which aggregates and displays the pathways of multiple event sequences through different states, summarizes each pathways outcomes, and allows the user to explore external factors that are related to pathway transitions. Outflow, unlike decision flow, incorporates external factors that are related to the event sequence and may have influence on changes in the event sequence. VAiRoma, developed by Cho et al. [11], is a visual analytics system that helps users to make sense of events, places, times, and the relationships between them permitting users to compare, make connections, and externalize their findings all within the visual interface. VAiRoma uses a data-driven and visual analytics approach to construct a narrative on history using content from Wikipedia articles. Like previous work, we utilize interactive views to enable users to explore event sequences both temporally and spatially. Our work differs in our use of intervention models to cue users to events in temporal datasets and then links the changes in event sequences to external data sources.

When working with multidimensional data, it may be necessary to explore the relationships within each dimension and across dimensions. To address the issues related to exploring hierarchical relationships in temporal information, we have seen, in past work, that multifaceted browsing and hierarchical timelines can offer effective solutions. For example, Plaisant et al. [58] demonstrated with LifeLines how to use hierarchical timelines and line properties to illustrate relationships or significance. LifeLines adds a clear and accessible view of temporal and causal relationships and takes advantage of the users ability to visually analyze information rich displays and interact with them. Chittaro et al. [10] proposed and explored three solutions to increase users ability to define temporal patterns and relationships. They introduce and comment on a number of aspects that have to be considered in visualizing temporal information and their proposed visual vocabularies for temporal relationships. Bade et al. [5] introduced several interactive visualization techniques, and a time visualization and navigation technique which enables users to inspect temporal data at different levels of detail and abstraction. In their paper they presented a novel way of visualizing time-oriented data and visualization techniques that enhance the understanding of the characteristics high-dimensional data. Continuum, developed by Andre et al. [1], applies multifaceted browsing to enable the exploration of hierarchical relationships and multiple timelines to allow users to inspect and compare relationships. Relationships in temporal data can be represented and explored between events across periods, enabling users to view specified levels of detail of any facet of interest. In order to allow users to explore the hierarchy of frames and examine their relationships, Event Cueing implements geographical controls for hierarchy filtering and a dendrogram containing all frames and categorizes their timelines for inspection and comparison. Our system enables users to examine relationships not only across frame types, but to also examine the relationship between frames and events.

The use of geographical maps to visualize and interact with geo-tagged text information allows the user to quickly find locations of interest and filter data based on spatial locality. NewsViews [29] is an automated news visualization system that generates annotated interactive maps using news articles to support trend identification and data comparisons. By leveraging text mining to identify key concepts and locations discussed in articles, Gao et al. are able to create thematic maps supporting various analytic tasks. Gao et al. employ text mining and extraction techniques to identify relevant features of articles automatically and use these features to select from data variables and spatiotemporal "cases". STempo [57], developed by Peuquet et al., is an integrated computational-geovisual environment to demonstrate T-pattern analysis can enable the discovery of patterns in complex spatiotemporal data. Peuquet et al. apply T-pattern analysis to an RSS news dataset, featuring events from news reports about Yemen. In their work, they demonstrate how the analysis technique can find associations among different event types where relationships are not known in advance. Tomaszewski et al. [64] introduced SensePlace to both visually and computationally support analyst sensemaking with text artifacts through the identification and visual highlighting of named entities. Building upon their previous work, MacEachren et al. [47] presented SensePlace2, to support situational awareness for crisis events by applying geovisual analytics to Twitter data. Utilizing explicit and implicit geographic information from tweets, SensePlace2 enables analysts to understand place, time, and theme components of evolving situations. SensePlace2 is the integration of linked visual-computation methods and a place-time-entity conceptual framework, enhancing situational awareness and sense-making. Kim et al. [39] present tools for generating personalized spatial analogies, re-expressing contextual spatial measurements with measurements that are more recognizable to the user. Using an automated approach, Kim et al. can take a user's location and generate personalized spatial analogies for the target of discussion. Employing an interactive application, Kim et al. enabled users to get a better understanding of a distance or area than with traditional measurements in an article with no spatial analogy. Field and O'Brien [27] illustrated a framework for mapping the spatial context of social media text to visualize patterns and reveal them to the user. They develop two proof-of-concept maps that support collaborative real-time mapping and the organization and display of information for mass user events. Their framework, "cartoblography", enables the mapping of spatial context of micro-blogging to more effectively communicate patterns exhibited by data. Similar to previous works, we apply text mining and extraction techniques to extract spatial locality to visualize relevant data geographically helping users to grasp where events are occurring and where frames are dispersed. In our work, we use a geographical layout to add another dimension of filtering, to visually inspect the dispersion of frames and events across Africa, and to compare their frame and event distributions.

Spatiotemporal visual analytics has used techniques, such as anomaly detection, statistical models, and cluster analysis, to increase hypothesis generation and situational awareness [9, 50, 60, 63]. Chae et al. [9] introduced a visual analytics approach that combines seasonal trend decomposition together with traditional control chart methods to find unusual peaks and outliers. Their approach provides analysts with scalable and interactive social media data analysis and visualization to explore and examine abnormal events. Analysts can extract major topics and apply analtyic techniques to improve situational awareness in a highly interactive process. Malik et al. [50] presented a visual analytics framework that provided proactive and predictive environments to assist in understanding and decision-making through the application of statistical analysis. In order to provide analysts with predicted levels of future activity, Malik et al. apply seasonal trend decomposition in a spatiotemporal visual analytics context. Their approach enables analysts to to focus and hone in on appropriate geospatial and temporal resolution levels. Sips et al. [60] introduced a visual analytics approach that computes statistical values for time series, detects patterns and allows users to visually inspect and evaluate the patterns. In their work, they demonstrate how their approach allowed scientists to gain new insights across two scientific scenarios in numerical time series. The approach proposed by

Sips et al., captures characteristics of temporal behavior regardless of time scales and starting positions. ScatterBlogs, presented by Thom et al. [63], automatically detects spatiotemporal anomalies using cluster analysis and event differentiation to classify event candidates and examine them on a global scale. Thom et al.'s approach allows analysts to visually examine and analyze microblog messages on a geospatial view using scalable aggregation and geolocated text. Their techniques work in realtime using interactive analysis with automatic anomaly detection at local and global levels. Similar to past work, we use anomaly detection and statistical modeling to detect interventions and patterns of interest allowing the user to visually inspect these patterns and the data they represent. Our work differs in that we use external data sources to detect interventions and improve our statistical modeling.

2.3 Visual Analytics for Data Linkage and Filtering

Data linkage through visual exploration enables users to fact check, discover external factors, correlate domain properties, and determine predictive links [38, 49, 65, 72]. Factful was presented by Kim et al. [38], to provide users with fact-checking support in a a web-based annotative article reading interface leveraging the linkage of news articles containing budgetary information with open government data. Factful enhances news articles providing contextual budgetary information and fact-checking to support budgetary discussions online. By linking news articles and open government data, Factful is able to engage taxpayers in meaningful discussions around budgetary issues with rich context from budget facts. Feng et al. [65] developed a visual analytics interface that used secondary data linking to associate surnames to interactively compare distributions of names with regards to spatial location similarity and income.

In their work, they enable interactive similarity exploration linking location data to demographic data visualizing the spatial distributions of names and relevant data from secondary data sources. Users are able to explore the similarity between surnames, their income distributions, and their spatial distributions. Contextifier by Hullman et al. [36] automatically links news stories to stock prices through custom annotations cueing user in on important events that may have caused shifts in prices. Contextifier provides users with a stock timeline graph with custom annotations that reference to the content in a news article enhancing the financial decision making process. This data linking allows users to make more informed decisions and hypothesize why shifts in prices occurred. To enable users to discover correlations and potentially causal or predictive links, Malik et al. [49] utilized the Pearson's product-moment correlation coefficient to detect trends or periodic patterns. Their approach allows users to understand the underlying foundations of the analysis and interactively explore and analyze correlations among datasets. Users are able to observe patterns and identify regions of probable activity in a visual analytics environment. Lu et al. [44] described a visual analytics approach to generalize the integration of social media data to other domains and demonstrated its use in predictive analysis. Their visual analytics toolkit extracts data from Twitter and Bitly to predict movie revenue and ratings in an interactive environment. Li et al. [43] present a detailed overview of strategies for the integration of social media and remote sensing data in time-critical applications to improve emergency response. In their work, they explore the relationship between environmental phenomena and social media responses, illustrating the potential of using social media data as a complement to remote sensing data sets for emergency response applications. Similar to previous work, we use statistical analysis to determine casual relationships between datasets, but we also use external linking to detect outliers and cue users to events that may be of interest.

Cross-filtering and linked views provide analysts with tools to quickly navigate and explore multiple collections of large data sets simultaneously. MultiFacet, presented by Henry et al. [34], extended existing faceted browsing systems to allow users to navigate large collections of multimedia from the text and image content of the data. Their work addressed the limitations of browsing collections of images and video by allowing allowing users to browse based on their actual visual content. Through the construction of facets the data can be browsed using visual facets combined with text facets exposing underlying relationships. Weaver [67, 68] described a method for cross-filtering data across pairs of views and strategies for constructing coordinated multiple views for visual analysis. In his work, he described how cross-filtering can used across designs and customized across dimensions to support complex analytic processes and queries by employing simple interactions. To allow for the formulation of queries that simultaneously combine temporal, spatial, and topial data filters, VisGets was introduced by Dörk et al. [25]. VisGets is a system of interactive query visualizations of Web-based information that utilizes online information in a web browser. Through the use of three linked visualizations (temporal, spatial, and topical), users can explore complex queries combining filters from more than one data dimension allowing them to gain casual insight into large collections of Web resources. Later Dörk et al. [24] introduced a system with mutually linked views supporting cross-filtering along topics, participants, and time ranges to provide an overview of large-scale conversations on twitter. In our work we use coordinated multiple views to filter multiple data sources, link information and tasks, and to help the user generate a correct understanding [59].

2.4 Collaborative Visual Analysis

Tools that enable collaborative visual analysis are able to enhance decision making processes and increase sense-making [7, 13, 33, 73]. This analysis approach allows diverse groups of users to explore more information and gain better insight through collaborative interactions. Examples of this include the development of the sense.us Web application by Heer et al. [33] to explore the possibilities for asynchronous collaborative visualization across a variety of visualization types. The user studies by Heer et al. using sense.us showed that combining conversation and visual data analysis can help users gain a wider and deeper understanding. In their work, they describe mechanisms for asynchronous collaboration using information visualization as a means of sharing, discussing, and navigating data. Kraut et al. [40], show that having shared visual spaces helps collaborators to have faster and better task performance through a study where they disaggregated the features of a shared workspace. In their work, they examine and share their findings the impact of temporal delay in shared spaces on task performance and communication. In addition their work focuses on visual complexity and temporal dynamics in shared visual spaces and their importance for difference tasks. The Entity Workspace system by Bier et al. [7] was modified to support collaboration where their studies showed that the collaboration tools had a positive effect on collaboration and information sharing. From work on Entity Workspace, Bier et al. presented five design guidelines for collaborative intelligence analysis focusing on the organize of information and its visualization. Convertino et al. [13] developed a geo-collaboration software prototype, which showed through distinct sets of measures that the prototype supported both content and process common ground. From their prototype they introduce design rationale drawn from their empirical work and related research and the implications for teamwork in software environments. VisTiles, presented by Langner et al. [42], is a conceptual framework that enables users to explore multivariate datasets across multiple devices. VisTiles enables users to interact with coordinated and multiple views in an individualistic or team focused workflow. In their work, they demonstrate how users can benefit from combining devices and organizing them into layouts that suit the needs of their analytic tasks. Wu et al. [73] developed a collaborative sense-making system that was able to support teamwork through a multi-view, role-based design helping team members to analyze geospatial information, share and integrate critical information. Through the use of coordinated maps and activity visualizations, Wu et al. are able maintain group activity awareness and aid in the decision-making process. Their work highlights the potential to improve and extend collaborative tasks through informed design rationale, iterative design, and system evaluation. In our work we draw on techniques from previous works such as multiple views, asynchronous and synchronous collaboration to provide analysts with the ability to share their work and save their work for later. Analyst are able to track and tag key findings through their exploration increasing their ability to collaborate with others and support their hypothesis.

Chapter 3

SYSTEM DESIGN

The goal of our visual analytics framework is to enable the exploration of patterns, actors, and events over time to gain insight for predicting possible future unrest. 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

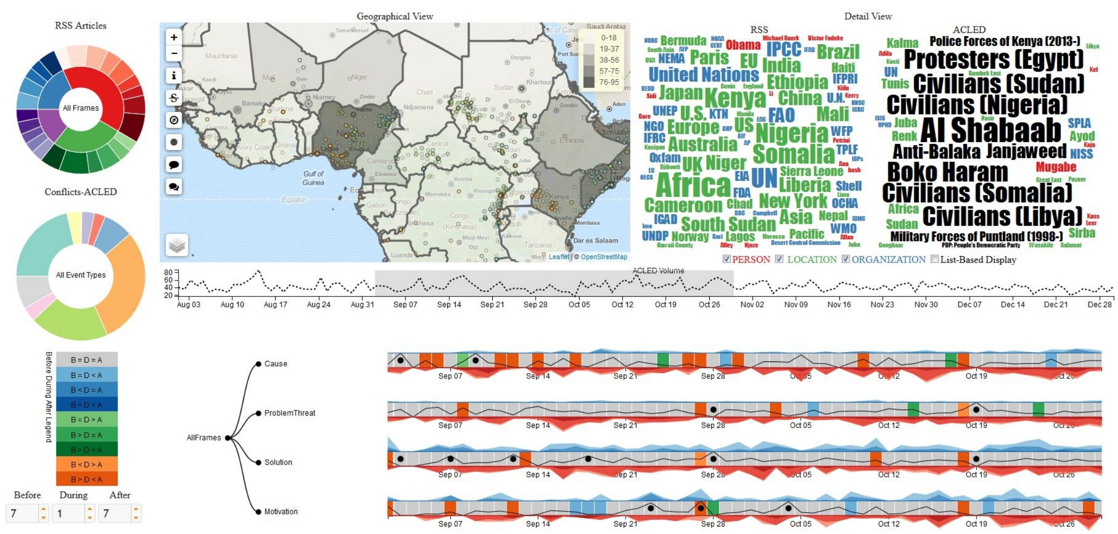

Figure 1. Overview of the event cueing visual analytics framework. The map view provides a geographical visual analytics environment to enable exploration of frames and entities over space and time. The detailed view to the right of the map switches between entity wordles and list-based displays. The 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.

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 Data Collection & Cleaning

The datasets used for the project are the Armed Conflict Location and Event Data Project (ACLED) [2] and a streaming RSS dataset.

ACLED: We sampled five months (August to December 2014) of political violence events from ACLED, which gave us approximately 6,500 events. Each event contains information pertaining to the day, parties involved, activity, location, sources and notes. These events are pulled from over 50 developing countries and include events that occur within civil wars and periods of instability. Each event is categorized into one of nine types defined in the ACLED Codebook [3]:

1. Remote violence - Remote violence refers to events in which the tool for engaging in conflict did not require the physical presence of the perpetrator. Remote violence notes that the main characteristic of an event is that a spatially removed group determines the time, place and victims of the attack. These include bombings, IED attacks, mortar and missile attacks, etc. Remote violence can be waged on both armed agents (e.g. an active rebel group; a military garrison) and civilians (e.g. a roadside bombing).
2. Battle-no change of territory - A battle between two violent armed groups where control of the contested location does not change. This is the indicated event type if the government controls an area, fights with rebels and wins; if rebels

control a location and maintain control after fighting with government forces; or if two militia groups are fighting. These battles are the most common activity and take place across a range of actors, including rebels, militias, and government forces, communal groups.

3. Battle-non-state actor overtakes territory - A battle where non-state actors win control of location. If, after fighting with another force, a non-state group acquires control, or if two non-state groups fight and the group that did not begin with control acquires it, this is the correct code. There are few cases where opposition groups, other than rebels, acquire territory.

4. Battle-government regains territory - A battle in which the government regains control of a location. This event type is used solely for government reacquisition of control. A small number of events of this type include militias operating on behalf of the government to regain territory outside of areas of a government's direct control (for example, proxy militias in Somalia which hold territory independently but are allied with the Federal Government).

5. Headquarters or base established - A non-state group establishes a base or headquarters. This event is non-violent, and coded when a permanent or semipermanent base is established. There are few if any cases where opposition groups other than rebels acquire territory. These events are coded as one-sided events without a second actor involved.

6. Non-violent activity by a conflict actor - This event records activity by rebel groups/militia/governments that does not involve active fighting but is within the context of the war/dispute. For example: recruitment drives, incursions or rallies qualify for inclusion. It also records the location and date of peace talks and arrests of high-ranking officials. The inclusion of such events is limited, as

its purpose is to capture pivotal events within campaigns of political violence. The notes column contains information on the specifics of the event.

7. Riots/protests - A protest describes a non-violent, group public demonstration, often against a government institution. Rioting is a violent form of demonstration. These can be coded as one-sided events. All rioters and protesters are noted by generic terms (e.g. Protester (Country)), but if representing a group, the name of that group is recorded in the 'ally' column.
8. Violence against civilians - Violence against civilians occurs when any armed/violent group attacks civilians. By definition, civilians are unarmed and not engaged in political violence, Rebels, governments, militias, rioters can all commit violence against civilians.
9. Non-violent transfer of activity - This event describes situations in which rebels or governments acquire control of a location without engaging in a violent act.

RSS News: The RSS News dataset is a collection of RSS feeds for the period of August to December 2014 provided by the Hugh Downs School of Human Communication. The media dataset is composed of RSS feeds from 122 English language news outlets in the Niger basin countries. RSS feeds were scanned hourly and filtered for relevance in a two-stage process. First, text from news outlets 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.

3.2 Data Processing

ACLED: Data processing applied to the ACLED dataset include entity extraction, sentiment extraction, and country indexing. For entity extraction, entities such as a person's name, locations, and organization are extracted from the dataset using the natural language processing tool CoreNLP [51]. For the ACLED dataset notes that are contained in each record, 367 persons, 998 locations, and 286 organizations were extracted. Sentiment extraction was also applied to the notes section of each ACLED record to get an overview of the attitude the source's author might have about the story. Three sentiment analysis classifiers are applied at the per sentence level. The three sentiment analysis classifiers used are SentiWordNet 3.0 [4], Stanford CoreNLP (Sentiment Analysis) [51], and SentiStrength [62]. Baccianella et al. [4] presented SentiWordNet, a lexical resource designed for supporting sentiment classification and oppinin mining applications. CoreNLP designed by Manning et al. [51] is an extensible pipeline providing natural language analysis from tokenization throught to coreference resolution. Thelwall et al. [62] introduced SentiStrength, which is an algorithm designed to extract sentiment strength from informal English text, using new methods optimized for the web. The three described analysis classifiers are used in a voting scheme to determine the sentiment of each sentence. Details on the voting

```
{
    "_id" : ObjectId("54f6150a341f9548d456a7a0"),
    "ACTOR1" : "GIA: Armed Islamic Group",
    "ACTOR2" : "Civilians (Algeria)",
    "YEAR" : 1997,
    "GWNO" : 615,
    "LATITUDE" : 36.766,
    "timestamp" : 852422400000,
    "SOURCE" : "www.algeria-watch.org",
    "Date" : "1997-01-05T00:00:00.000Z",
    "coordinates" : {
        "type" : "Point",
        "coordinates" : [
            3.05,
            36.766
        ]
    },
    "NOTES" : "7 January: Explosion of a bomb in the Didouche Mourad street in Algiers: 20 dead.",
    "TIME_PRECISION" : 1,
    "EVENT_TYPE" : "Remote violence",
    "INTERACTION" : 27,
    "GEO_PRECIS" : 1,
    "FATALITIES" : 20,
    "COUNTRY" : "Algeria",
    "INTER1" : "2",
    "ADMIN1" : "Alger",
    "ADMIN2" : "Bouzareah",
    "INTER2" : "7",
    "ADMIN3" : "",
    "EVENT_ID_CNTY" : "4ALG",
    "LOCATION" : "Algiers",
    "EVENT_ID_NO_CNTY" : 4,
    "LONGITUDE" : 3.05,
    "ALLY_ACTOR_1" : "",
    "ALLY_ACTOR_2" : "",
    "sentiment" : {
        "coreNLPScore" : 1,
        "sentiStrengthScore" : -2,
        "sentiWordNetScore" : -0.1104008192901471,
        "votedscore" : -1.333333333333333,
        "entropy" : 0.5991464547107982
    },
    "countryIndex" : 0
}
```


Figure 2. Example of a processed ACLED record which has been through the process of sentiment, entity, and country extraction. Each record is made up of a uniquely generated id, the named actor or actors involved in the event, the year, the geographic coordinates (lat,lng), the country, the date and timestamp of the event, the type of event, the notes detailing the event, the precision of the date and location information, the number of fatalities, the event source, the largest and second largest sub-national administrative region in which the event took place, and the sentiment scores.

scheme are provided in Section 3.7. An example of the resulting event data record can be found in Figure 2.

RSS News: Data processing applied to the RSS News dataset include entity extraction, geo-location extraction, frame encoding, sentiment extraction, and country indexing. The entity extraction is the same process as performed on the ACLED dataset and returned 19,756 entities which breakdown to 2,107 persons, 5,791 locations, and 3,146 organizations extracted from the RSS media dataset. Geo-location extraction was performed on each article based on where it was posted or the region the article discusses. Since this information may not be easily identifiable, the Data Science Toolkit was used to extract and geocode the articles [16]. Frame encoding was performed using trained coders. The trained coders coded each sentence into one


```
{
	"frame" : "Insufficient",
	"geocoding" : [
		{
			"name" : "Kenya",
			"coord" : {
				"lat" : 1,
				"lng" : 38
			},
			"type" : "country"
		}
	],
	"ner" : [
		{
			"mentionSpan" : "World Food Programme",
			"namedEntity" : "ORGANIZATION"
		},
		{
			"mentionSpan" : "Kenya",
			"namedEntity" : "LOCATION"
		},
		{
			"mentionSpan" : "United Nations World Food Programme",
			"namedEntity" : "ORGANIZATION"
		}
	],
	"sentence" : "World Food Programme to cut food rations to refugees in Kenya by half due to lack of funds camps
will receive reduced rations from the United Nations World Food Programme -LRB- WFP -RRB- as a result of insufficient
funding .",
	"sentiment" : {
		"coreNLPScore" : 1,
		"sentiStrengthScore" : -1,
		"sentiWordNetScore" : -0.7056403717338472
	},
	"sentimentflag" : 1,
	"geoCoord" : [
		{
			"name" : "Kenya",
			"country" : "Kenya",
			"coord" : {
				"lat" : 1,
				"lng" : 38
			}
		}
	] ,
	"countryIndex" : [
		64
	]
}
```


Figure 3. Example of a processed RSS sentence record which has been through the process of sentiment, entity, geo-location, country extraction, and frame encoding. Articles are broken up into sentences for sentence level analysis. Each record conatins a uniquely generated id, the frame type, the geographic coordinates (lat,lng), the country, the entities mentioned, the sentence, and the sentiment scores.

(or none) of 25 categories comprising four classes (cause, problems/threat, solution, motivation) that represent different types of framing for climate change. The 25 categories are:

- Causes of climate change
  1. CauseHuman - Human activity
  2. CauseNatural - Natural variation
  3. CauseUncertainty - Uncertainty
  4. CausePolicy - Policy causes
  5. Insufficient - Insufficient action
  6. OtherHumanFactors - Human disruption to mitigate climate change impact
- Threats/Problems
  7. ProbThreatEnvironmental - Problems and threats to the environment
  8. ProbThreatHealth - Problems and threats to Public Health
  9. ProbThreatEconomic - Problems and threats to economic consequences
  10. ProbThreatFood - Problems and threats to food security
  11. ProbThreatWater - Problems and threats to water security
  12. ProbThreatSecurity - Problems and threats to National Security
  13. ProbThreatSocialUnrest - Problems and threats to Social Unrest
  14. ProbThreatGenMultiple - General and multiple problems and threats
- Solutions
  15. SolutionConservation - Conservation as solution
  16. SolutionEducation - Education as solution
  17. SolutionInvest - Investment as solution
  18. SolutionInfraDev - Infrastructure and development as solution

19. SolutionPolicyProgram - Policy and program as solution
20. SolutionGoal - Goals of solution

- Motivation

21. MotivSeverity - Severity as motivation
22. MotivUrgency - Urgency as motivation
23. MotivEfficacy - Efficacy as motivation
24. MotivPropriety - Propriety as motivation
25. MotivGeneral - General motivation

After coding was completed 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 [41]. The sentiment extraction process that was performed on the ACLED dataset was performed on the RSS News dataset for the news text of each article. An example can be found in Figure 3 of the resulting data record.

3.3 Data Filtering

The visualization supports data filtering based on several user selectable properties: time period of evaluation, frame type, event type, and geo-location.

1) Time period: The time period for which data is show is controlled through a timeline with brushing for the periods of start and finish. The timeline also shows the volume of ACLED events per day.
2) Frame type: Each sentence of the RSS dataset is coded by a trained coder into one (or none) of twenty five categories comprising four classes (cause, problem/threat,

solution, motivation) that represent the different types of framing for climate change. The frame type selection is controlled through a donut chart that updates based on the distribution of frames.

3) Event type: Each political violence event is assigned one of nine event types as provided by ACLED. The event type selection is controlled through a donut chart that updates based on the distribution of events.
4) Location: Using a shapefile to define the countries of Africa, data can be filtered based on country borders. Clicking within a region outlined by the shape file updates the detailed view to display information within the region.

3.4 Geographical View

The geographical view is necessary to explore the ACLED and RSS datasets in terms of location. The ACLED and RSS datasets both contain geo-location information, this is processed and plotted to the map based on user selected inputs and modes of operation.

3.4.1 Exploring the RSS Dataset

In order to explore frames spatially, we created a spatial data distribution visualization. Past work focused on exploring frames through word cloud and lingoscope [19, 20, 21], however we wanted the user to be able to see where frames were located and how framing changed for those countries and regions. RSS frames are shown using the following views:

1. To show the total density of all frames for a given dataset, we use a choropleth that is colored sequentially based on the density of frames in each country.

Figure 4. The choropleth view shows the default choropleth map, which colors each country based on the density of all frames. The darker the hue of a country the greater it's density is relative to all other countries.

2. To show which frames had the greatest weight per country based on a user adjustable scale we used a weighted choropleth view. Each color corresponds to the appropriate framing category of either cause, problemthreat, solution, or motivation.

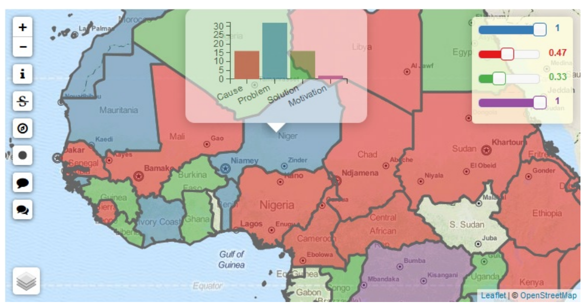

Figure 5. The weighted category choropleth view shows a weighted choropleth map, which colors each country based on the weighted frame density. The dominant frame for each country is shown and this changes as the user adjusts the weights which are in the upper right hand corner of the map. The bar chart shows the numerical values for each frame type per country.

3. To show the overall collection of frames for each region, based on a set radius a pie chart is drawn on the map. The pie chart has a breakdown of each frame type for the region it defines and changes based on zoom level. By supporting different zoom levels we allow for the user to explore the data from a wide range of views.

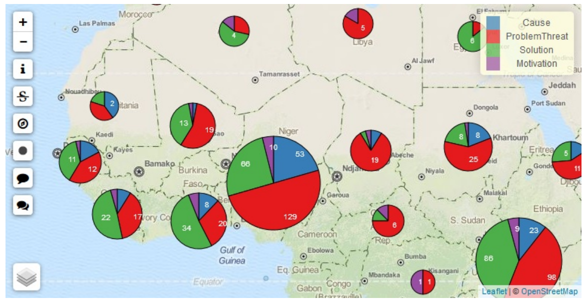

Figure 6. The cluster pie view shows pie glyphs on the map displaying the proportional distribution of different frame categories in each cluster. The size of a the glyph represents the total number of frames relative to the other regions.

3.4.2 Exploring the ACLED Dataset

In order to explore the political violence events from the ACLED dataset spatially we use the geographical view to plot markers that are colored based on their event type. This allows the user to visually examine where events are occurring and to interact with them spatially rather than through their text sources.

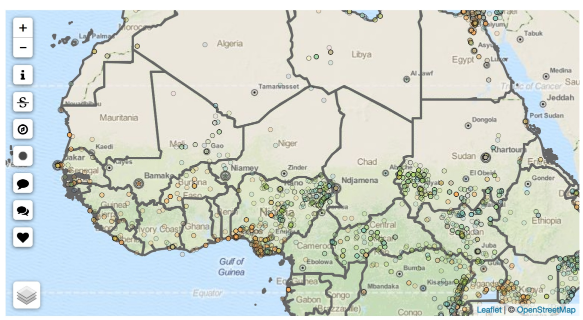

Figure 7. ACLED Event Markers (Color): Remote Violence (Gray), Battle-No change of territory (Turquoise), Battle-Non-state actor overtakes territory (Yellow), Battle-Government regains territory (Purple), Headquarters or base established (Red), Non-violent activity by a conflict actor (Teal), Riots/Protests (Orange), Violence against civilians (Green), and Non-violent transfer of territory (Pink).

3.4.3 Exploring Happiness

In order to explore the happiness from the RSS dataset spatially, we use a choropleth map. The choropleth map uses a divergent color scheme from happy to sad, with neutral being the center. Past work focused on exploring happiness temporally or geographically using tweet data [23, 52] however, we wanted the user to be able to explore happiness spatially and temporally. In our visualization we map our own sentiment extraction techniques and compare our results with LabMT [52] using a RSS news dataset. The Language Assessment by Mechanical Turk (LabMT) word list consists of about 10,000 individually scored words from four text sources: Google Books, music lyrics, the New York Times and Twitter [52]. The LabMT word list was applied to the RSS news data to measure sentiment in addition to our sentiment analysis classifiers.

1. To show the well-being for each region within Africa, we use a choropleth that is colored sequentially based on the average sentiment within each country.

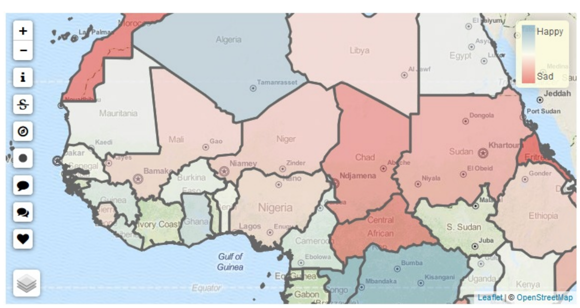

Figure 8. The sentiment happiness view shows a choropleth map where each country is colored based on its relative happiness. The happiness in this view is determined by our sentiment analysis techniques. Blue colors represent happier countries, white colors represent neutral countries, and red colors represent sad countries.

2. To show the well-being for each region within Africa, we use a choropleth that is colored sequentially based on the average LabMT score within each country.

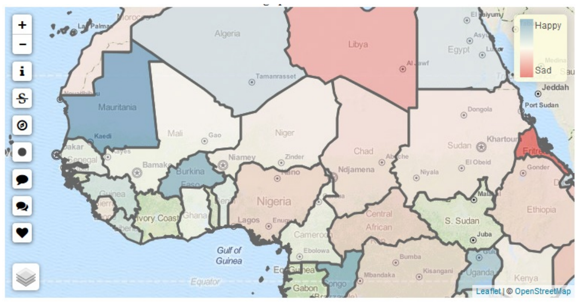

Figure 9. The LabMT Happiness view shows a choropleth map where each country is colored based on its relative happiness. The happiness in this view is determined by the LabMT score. Blue colors represent happier countries, white colors represent neutral countries, and red colors represent sad countries.

3.4.4 Entity Lens

In order to explore the specific entities involved in the news articles and political events spatially we designed a text lens. Entity extraction is performed on both the ACLED and RSS datasets allowing us to explore entities based on the geo-location information that the entities are tagged to. The types of entities extracted are persons' names, locations, and organizations, as well as a set of predefined entities that a user may wish to explore. The lens maps the most frequent words closest to lens's circumference based on available drawing space. The font size of an entity is relative to its frequency within the highlighted area. The lens updates the text displayed based on the size of lens drawn, the area it's drawn to cover, and any areas it's moved to. The Entity Lens operates in three different user switchable modes:

1. RSS Mode: In this mode the lens will only display entities extracted from the RSS dataset. Information displayed is dependent on the user's selection of frame types.

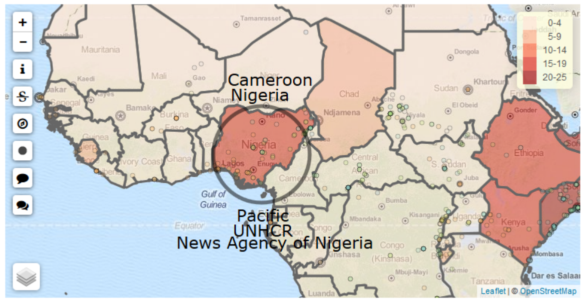

Figure 10. The RSS entity lens on the map shows the most frequently appearing named entities from the RSS news dataset. The area under the lens is used to filter the geocoordinates of all RSS news articles and only returns the entities of those that within the lens.

2. ACLED Mode: In this mode the lens will only display entities extracted from the ACLED dataset. Information displayed is dependent on the user's selection of event types.

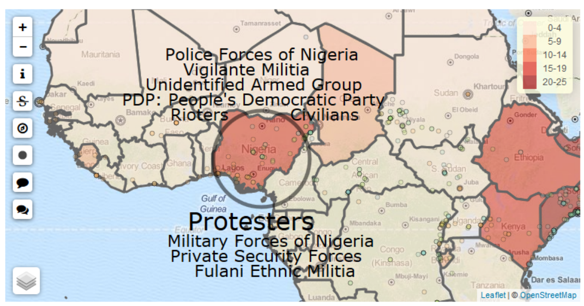

Figure 11. The ACLED entity lens on the map shows the most frequently appearing actors from the ACLED dataset. The area under the lens is used to filter the geocoordinates of all ACLED events and only returns the actors of those that within the lens.

3. Comparison Mode: In this mode the lens will show a combination of entities extracted from both the RSS dataset and the ACLED dataset for comparing. On the left side of the lens RSS news' entities are displayed and on the right side ACLED entities are displayed. Information displayed is dependent on the user's combined selection of frame and event types.

Figure 12. The comparison entity lens on the map shows the most frequently appearing actors from the ACLED dataset and the most frequently appearing named entities from the RSS news dataset. The area under the lens is used to filter the geocoordinates of all ACLED events and RSS news articles. After filtering the articles and extracting the relavent information from each dataset the RSS news' entities are shown on the left and ACLED events' actors are shown on the right.

3.5 Detail View

The detail view allows the user to explore the entities for both datasets, the RSS article titles and sentence sentiment, and the ACLED notes. The information displayed changes based on a user selected period, subset data selections, and country selection. The detail view operates in two different modes:

1. Entity Wordle: in this mode the detail view shows the most frequently named entities for both the RSS and ACLED datasets. The entities are broken into three classes: Person, Location, and Organization. The user can choose which classes to be display for both datasets. For the ACLED wordle there is a fourth class which denotes the actors involved in political violence events extracted from the dataset.

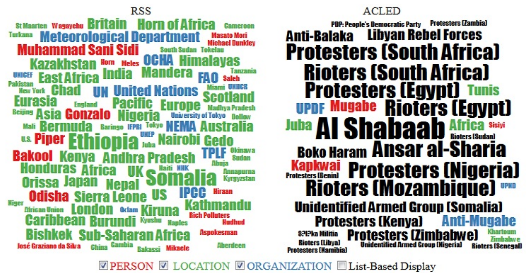

Figure 13. The entity wordle view is the entity wordle display in which users can choose three classes of entities (person, location, organization) to show. Black text in the ACLED Wordle indicates an actor in the events. For both views: Red text represents an entity of the class person, green represents an entity of the class location, and blue represents an entity of the class organization.

2. List-based Summary: in this mode the detail view shows a summary of the text information from the ACLED events and RSS news articles. The RSS details include the date of the article, the title, and the frames identified within the article. The frames identified in the articles are shown as colored squares. Each square represents a framed sentence and is colored by its frame class's color. The ACLED details include the date of the event, the notes from the event, and is colored based on the type of event.

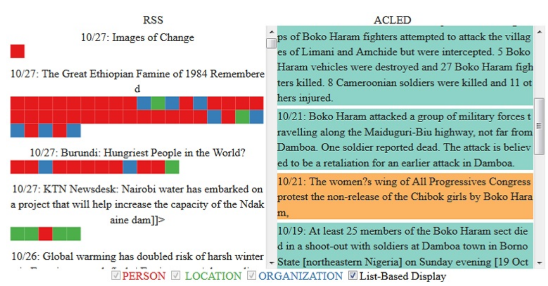

Figure 14. The list-based summary view displays the title of media articles and the summary information of the secondary dataset are in order of time. The frame information of each article is summarized into colored squares (the color of the square matches the frame class) in the sentence order from the article. In this example, the ACLED event notes are filtered by clicking on the entity text 'Boko Haram'.

3.6 Hierarchical Frames Timeline View

In order to cue users in on events that are statistically interesting within the data, we provide a timeline view (Figure 15). The timeline view uses two time series analysis models to visualize possible interventions. Enabling the user to have direction within their exploration permits them to find events and correlations of interest without performing prolonged exploration.

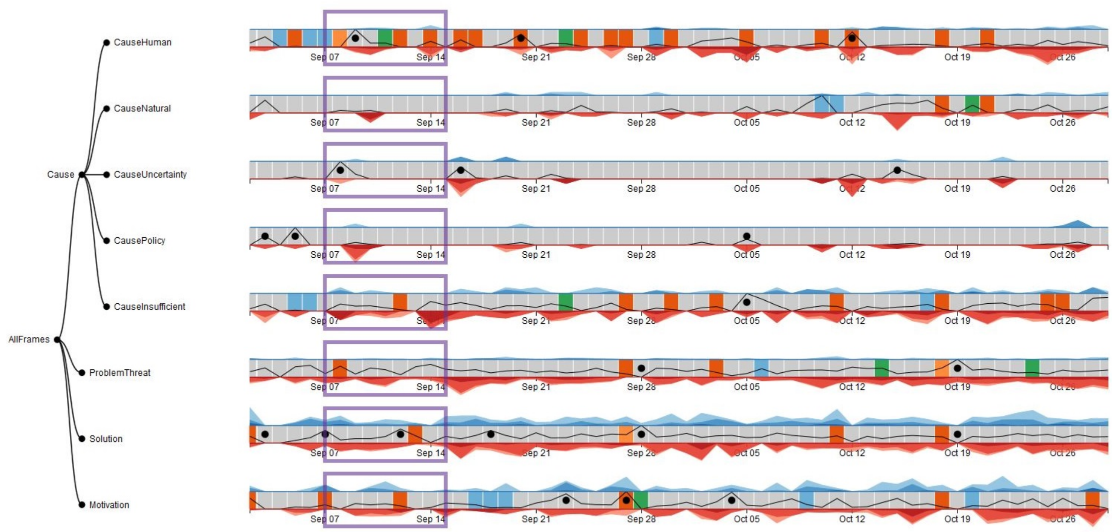

Figure 15. Hierarchical timeline view showing intervention modeling results, Before-During-After analysis results and sentiment river for each expanded frame or frame category are shown. The frame structure is displayed as a dendrogram on the left. Clicking on the node can expand/collapse its children. The timeline associated with each leaf node is shown on the right.

Each timeline shows the relative volume of frames per day. This 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. Each document has a number of sentences that are encoded with a single frame, and the percent of framing for a document is the number of sentences within the document associated with a given frame divided by the total number of sentences framed.

In addition to intervention detection, sentiment analysis is applied to the RSS data giving each timeline a two-sided uncertainty-based stack river. The timelines can be expanded from four high level classifications (cause, problemthreat, solution, motivation) to twenty five categories that make up those classifications.

3.6.1 Intervention Modeling

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:

$$y_t = v(B)x_t + N_t \tag{3.1}$$

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 [53]. 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:

$$y_t = v(B)\epsilon_t + \omega I_T^{(t^*)}, \text{ where } I_t^{(t^*)} = \begin{cases} 1 \text{ if } & t = t^* \\ 0 \text{ if } & t \neq t^* \end{cases} \tag{3.2}$$

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 thesis, intervention models were used to detect outliers, i.e. cues to events that may be of interest to an analyst, for each of the 25 frames over the time period of August 2 to December 31, 2014. Initial analysis of each time series (each frame) indicated that there was no significant autocorrelation present. Therefore, the intervention model can be simplified to:

$$y_t = \mu + \omega I_t^{(t^*)} + \epsilon_t \tag{3.3}$$

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 [53]. 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 an 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 3.2 would be used in such cases.

3.6.2 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* [69], 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 [54],

$$\alpha = \frac{\alpha_F}{\#test}, \tag{3.4}$$

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 Figure 16. The 9 types are visualized in different color blocks on the time line for each frame, as shown in Figure 15 and Figure 17.

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)

Figure 16. The pattern summary for Before-During-After analysis. Each pattern is associated with a unique hue as shown in the lower left-hand legend of Figure 1. The user can enable and disable different patterns depending on their analysis.

The statistical tests' results are visualized in our timeline view for each frame and frame categories, shown in Figure 15 and Figure 17. 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, an analyst can adjust the before, after and intervention (during) periods using the controls seen in Figure 1 (lower left).

3.7 Frame Sentiment Analysis

In order to visualize the relation of the underlying sentiment of the media to its framing a sentiment analysis has been employed. A novel entropy-based sentiment river is shown to reveal the uncertainty of sentiment over time. An ensemble voting scheme from multiple classifiers is used to determine the final sentiment label. From previous work [10], uncertainty is defined as:

$$UC = 1 - normalize(Max\_Entropy - Entropy) \tag{3.5}$$

Where Entropy is defined as follows:

$$Entropy = -\sum_{i=1}^{K} \frac{V(y_i)}{C} \log \frac{V(y_i)}{C} \tag{3.6}$$

Here V(y_i) is the number of "votes" that a class (K_i) receives from the committee members' predictions, K denotes the number of classes, and C is the committee size. Max_Entropy is the highest possible entropy given C and K. A stacked area and low uncertain area is stacked as shown in Figure 17 to explore the change of certainty over the media stream, the volume of certain, and uncertain sentiment labels.

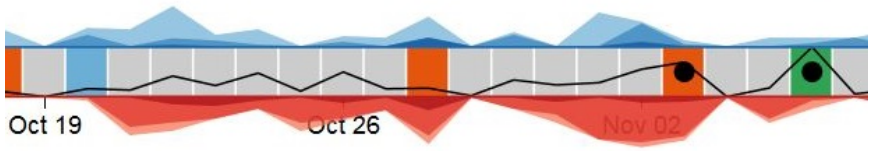

Figure 17. Sentiment stacked area chart on bi-side of the time series view. The blue area represents positive sentiment and the red area river represents negative sentiment. The darker the area color is the more certain the label is for those sentences' sentiment class.

In previous work by Lu et al.[45], 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 17 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.

Once the user has selected a period of interest for exploration the system is able determine if a relationship of predictive causality exists between the two data sets. The system uses the Granger Causality Test to determine if ACLED events Granger-cause RSS news articles. By definition a variable X Granger-causes Y if Y can be better predicted using the histories of both X and Y than it can using the history of Y alone. Causality is tested using the following causal model provided by Granger et al. [31]:

$$\begin{aligned} X_t &= \sum_{j=1}^{m} a_j X_{t-j} + \sum_{j=1}^{m} b_j Y_{t-j} + \varepsilon_t, \\ Y_t &= \sum_{j=1}^{m} c_j X_{t-j} + \sum_{j=1}^{m} d_j Y_{t-j} + \eta_t \end{aligned} \tag{3.7}$$

Where X_t and Y_t are stationary time series, ε_t, η_t are taken to be two uncorrelated white-noise series, i.e., $E[\varepsilon_t \varepsilon_s] = 0 = E[\eta_t \eta_s], s \neq t$, $E[\varepsilon_t \varepsilon_s] = 0$ all t, s, and m is an integer in [1,t]. The definition of causality given above implies that X_t is causing Y_t if some c_j is not zero.

Based on if ACLED events are found to Granger-cause RSS news articles the intervention model will produce better results when predicting intervention points. This also tells the user that the events within the selected time period contain useful information related to the articles. If ACLED events are found to not Granger-cause RSS news articles then they may not be useful in predicting changes in frames. Since we are only concerned with the one way causal relationship between ACLED events and RSS articles we can simplify the linear regression model to the following [31]:

$$\begin{aligned} Y_t &= a_0 + \sum_{j=1}^{m} a_j Y_{t-j} + \varepsilon_t, \\ Y_t &= a_0 + \sum_{j=1}^{m} a_j Y_{t-j} + \sum_{j=1}^{m} b_j X_{t-j} + \varepsilon_t \end{aligned} \tag{3.8}$$

Where ε_t is taken to be an uncorrelated white-noise series, i.e., $E[\varepsilon_t \varepsilon_s] = 0 = E[\eta_t \eta_s], s \neq t$, and m is an integer in [1,t]. If the prediction variance is statistically significantly improved in the second model we can determine that the events identified in the dataset seem to be driving the media stream. This method should only be considered as "predictive causality" due to the lack of field testing and experimentation that "true causality" requires [22].

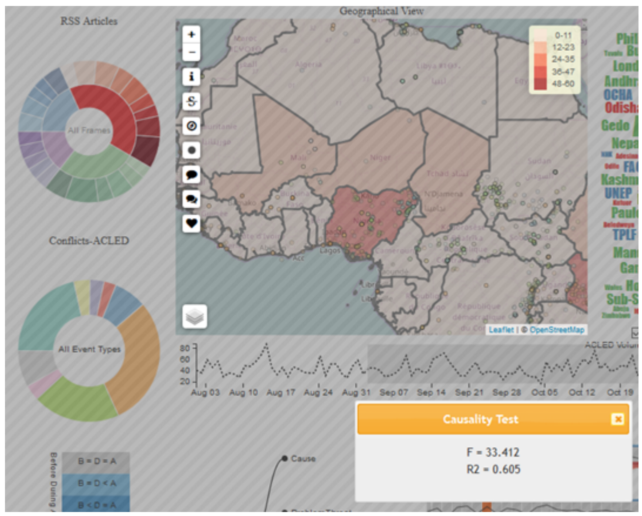

Figure 18. The causality test allows users to select a time frame for which to test an RSS Article frame type and ACLED Event Types to determine if predictive causality exists. If the prediction variance is statistically significantly improved in the second model we can determine that the events identified in the dataset seem to be driving the media stream.

This analysis can still be a useful indicator for analysts and help them determine trends and drivers of interest. Figure 18 shows the results of running a causality test on the RSS Problem/Threat frames for the period of September 1st through October 31st with all event types selected.

3.9 Collaboration View

The collaboration view allows the user to manage their findings and the findings of other users within their space. They can add, edit, and delete any of the comments or tags made within their space and merge their findings with others. In this view the user can also choose to join another user's space for coordinated analysis or allow other users to join their space. User spaces are shared through their unique id, making it quick and easy to manage and join sessions. This view also allows users to save, view, edit, and load their configurations. At any point a user can save the current state of the system and return to it later. The user is prompted when loading or saving configurations to tag them with the appropriate metadata information for managing their explorations.

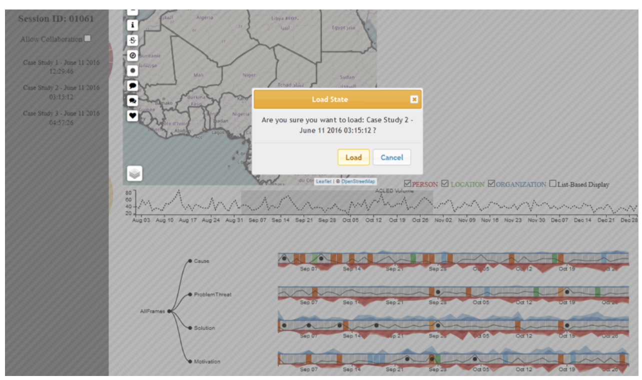

Figure 19. The configuration view allows users to save, view, edit, and load their configurations. In this image the user is selecting to load the configuration data from a past exploration. After loading the configuration data, all views will be updated to their state that was stored in the configuration data.

Chapter 4

CASE STUDY

Our visual analytics framework uses real conflict data in combination with RSS data to examine when and how framing about a topic has shifted. The system's entity extraction and sentiment analysis act as guides through data exploration and help users to focus on themes within those topics. Through the use of intervention models for time series analysis, the system can highlight statistically significant intervention points to cue users to time periods of interest. After finding a period of interest, its articles and event data can be examined geographically to understand the spatial distribution of frames. The text information from before, during, and after this period can be examined directly or through entity extraction to determine the parties involved and the locations and organizations of greatest interest. Combining these techniques allows the system to support spatiotemporal analysis of frames and cue users when frame changes have occurred.

In this section, I will demonstrate how our visual analytics tool can help analysts to explore and analyze large datasets 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 [55]. 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 20. 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 20) 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 20(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.

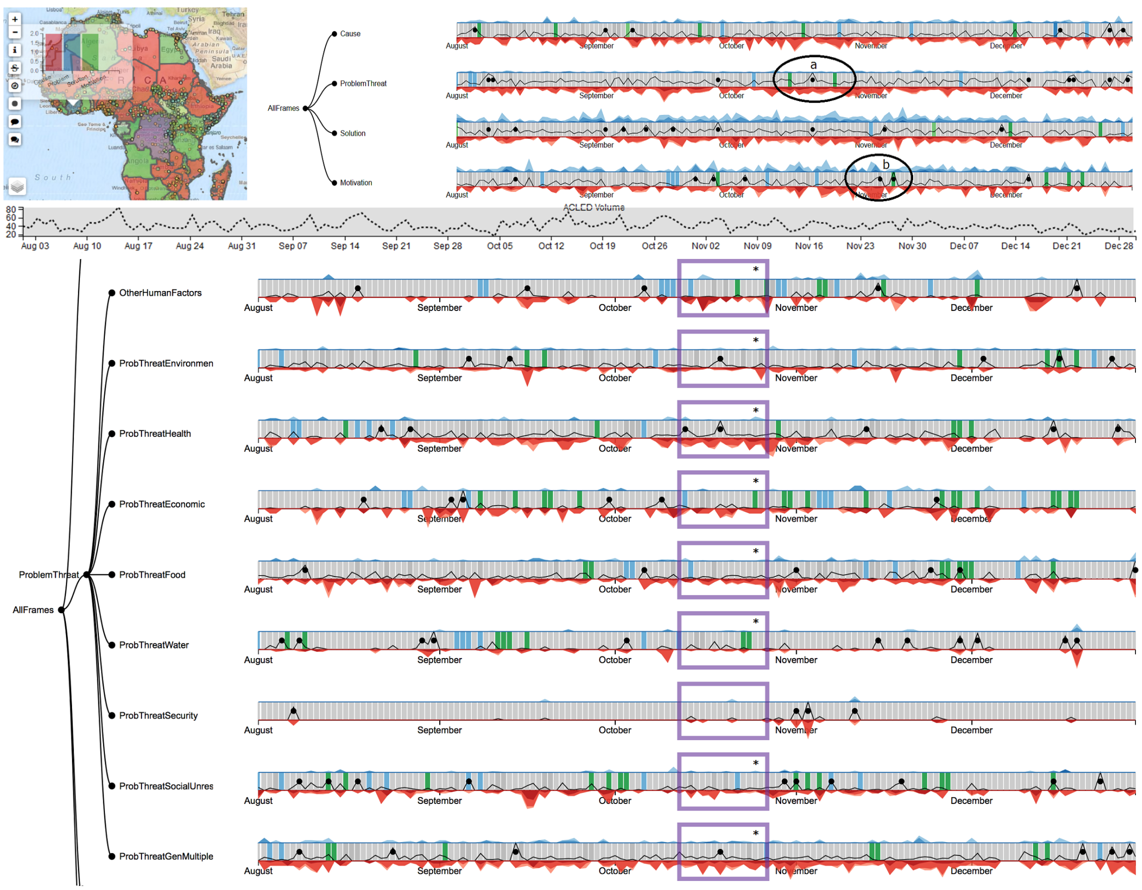

Figure 20. Exploring the RSS news dataset spatially and temporally for the entire time period that there is data available. The spatial map shows a weighted choropleth map with all frame class equally weighted. The add-on histogram shows the frame volume and distribution of the Republic of Côte d'Ivoire. 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.

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 21(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. A causality test is performed and the analyst finds a best fit model with a $lag = 2$, $R2 \approx 0.54$ and $p-value \approx 0.25$. Although the R2 is low the ACLED events under analysis may still contain drivers of the changes in framing for the period giving cause for further investigation. The analyst turns to 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

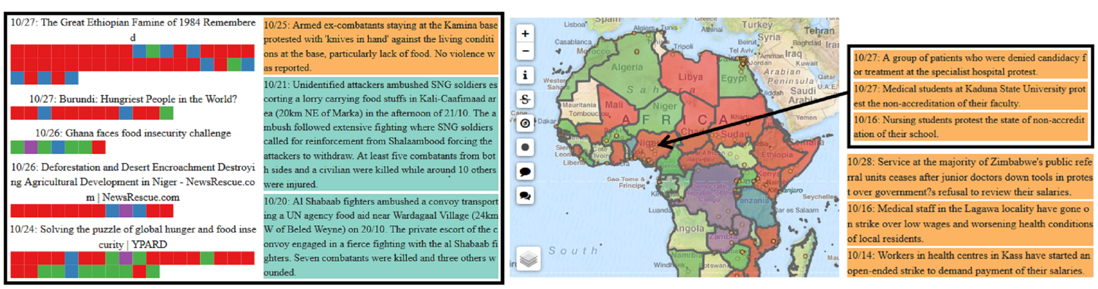

Figure 21. The geographical and detail view for exploring RSS news and ACLED data. This figure shows the analyzing time from October 11th to October 28th. The geographical view colors each country by the majority frame class and displays riots (orange dots) to represent the ACLED events. The detail view 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.

discourse in the hopes to alleviate concerns 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 Figure 20 (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 22. 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).

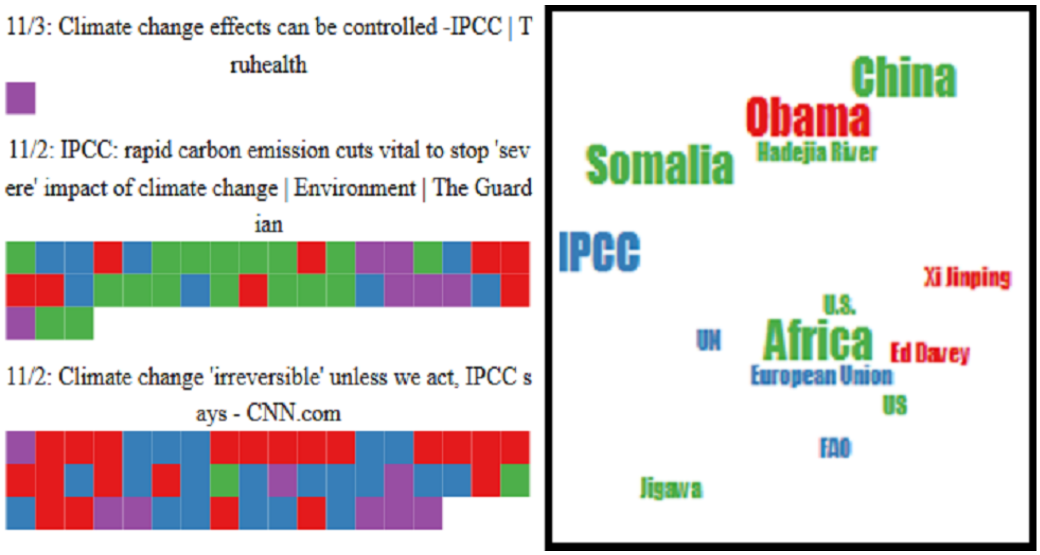

Figure 22. Example RSS articles and the entity wordle for the time period of Oct. 28th to Nov. 11th exploring motivation frame. The left side article summaries show examples of news report relating to the IPCC and the right side wordle emphasizes the most frequent entities appearing in those articles, such as IPCC, Obama, and China.

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.

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 Exploring Happiness in Africa

In the previous case studies, we discussed how analysts were able to explore problem and motivation frames in Africa and their preliminary findings. Another use of our system is the exploration of the emotion, happiness or sadness, which the media is dispersing across the region. In our previous case study, we highlighted the Intergovernmental Panel on Climate Change (IPCC) on November 2nd and the noticeable spike in positive sentiment values between November 9th and November 12th. We can explore the periods leading up to the panel and the periods after to see if there is noticeable change in the emotions across the region.

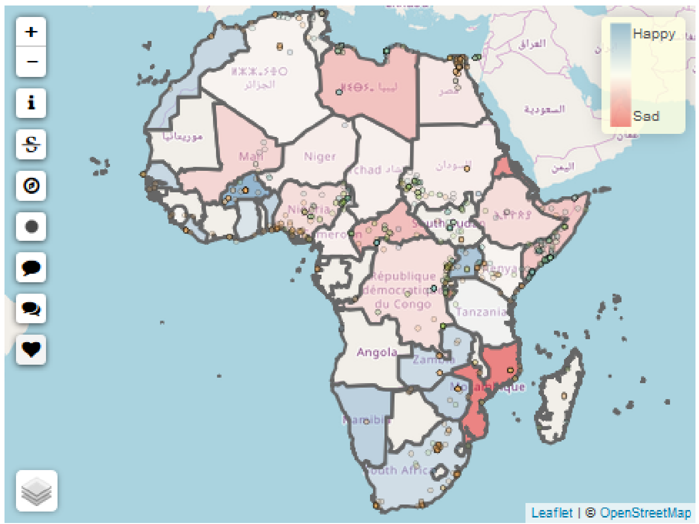

Figure 23. The geographical view for exploring RSS news and ACLED data. This figure shows the analyzing time from October 4th through October 31st. The geographical view colors each country by the overall emotion found in news articles and displays dots to represent the ACLED events. From the geographical view, we can see there is a mix of emotions that can be detected from the articles collected during this period.

Before our inspection, we would hypothesize that this positive spike would show across all regions that released articles in the period after the panel and that the map should show more happy or positive outcomes in media discussions. This an outcome that could be expected due to nature of the panel and its push for change and cooperation across and between countries to address climate change. For this exploration we will be looking at all frame and event categories to see if there is an overall shift, keeping in mind that the ACLED events during these two periods play a role in the emotions we will examine. We can first begin by examining press coverage before the conference from October 4th through October 31st. We can see in the map view of Figure 23, that there is a mix of emotions that can be detected from the articles that were written during the period in all countries where articles are available.

Turning our focus to the period after the conference, November 3rd through November 30th, in Figure 24 we can see an overall shift towards more neutral and in some cases happier news that is being released. In particular, we can see shifts in the northern countries of the region that change from a mildly sad tone to an almost mild happy tone. For articles released from countries in the eastern and western parts of the region shift from a mildly sad tone to a more neutral tone.

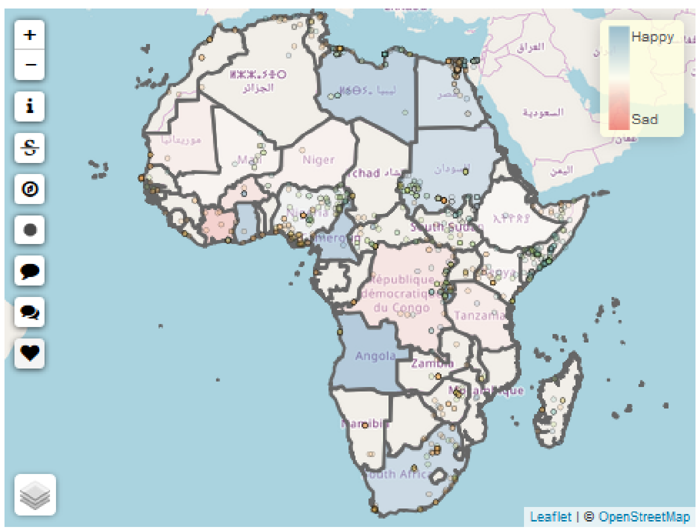

Figure 24. The geographical view for exploring RSS news and ACLED data. This figure shows the analyzing time from November 3rd through November 30th. The geographical view colors each country by the overall emotion found in news articles and displays dots to represent the ACLED events. From the geographical view, we can see there is a shift towards an overall happier perspective that can be detected from the articles collected during this period.

If we look into the RSS articles for this period, we can see there are more articles that have an overall happier outlook. Many of these articles contain sentences mentioning solutions and motivations. Some of the headlines from these articles are:

- "US Agreement With China Paves The Way For A Global Climate Deal"
- "Climate change as an opportunity for the poor"
- "U.S. and China reach historic climate change agreement"
- "White House, China agree to greenhouse emissions deal"'
- "US and China leaders in 'historic' greenhouse gas emissions pledge"

From these results, we can see that our hypothesis did hold true however, there are other factors at play that we can investigate. When looking at the ACLED events over this period we see there is a subtle decrease in events over this period, but not enough to substantiate this change in emotion without inspection of the conflicts. In general, the proportions of event types from the period after the panel to those of the period before the panel are similar, which gives us this indication.

4.4 Analyst Feedback

Our first two 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.

Chapter 5

CONCLUSIONS & FUTURE WORK

In this thesis, a visual analytics tool that can be used to analyze correlations between types of armed conflicts and rss news articles was presented. The visual analytics tool allows for the exploration of social climate and the visualization of its changes with respect to space and time. Using the tool it is possible to discover patterns in the ACLED dataset by linking the RSS dataset and using the coordinated views within the system. It is also possible to recognize correlations between conflict types and news articles and use them in coherence to explore complex hypothesis to link events together. These interactions are supported by the utilization of techniques from both the text and time series analysis domains. Leveraging these techniques users can be cued into time periods of interest and quickly explore the hierarchy of frames and their evolution. Although past work has been done in the development of tools for frame analysis [19, 20, 21], their support is limited to comparing corpora of text and topical terms within these text. Our framework expands on past work enabling sentiment analysis and intervention modeling which can provide different insights than previous works. This thesis has addressed the challenges of identifying framing in media, exploring how media is framed over space and time, determining the drivers of change in media framing, exploring and supporting complex hypotheses, effectively combining multiple data sources, and collaborative visual analysis.

In the case studies the capabilities of our system to explore and analyze large datasets by applying the methods described in this thesis has been demonstrated. In our first case study we presented the process of how an analyst can explore problem frames in Africa. In the second case study we described the process of exploring motivation frames in Africa. Both of these processes were supported by the system's entity extraction and sentiment analysis acting as guides through data exploration and helping users to focus on themes within their respective topics. The use of intervention models allowed the system to highlight statistically significant intervention points to cue users to time periods of interest. After finding periods of interest, the corresponding articles and event data could be examined geographically to understand the spatial distribution of frames. Information from before, during, and after the periods could be examined to determine the parties involved and the locations and organizations of interest.

Even though we are able to provide an environment for the exploration of these datasets in an interactive fashion. The system does not come without limitations. One limitation is in the causality analysis; currently users are limited to testing one frame type at a time. This limitation can hinder an analyst's ability to detect causality across multiple frame categories limiting their exploration process. Another limitation is that users cannot combine the models in their analysis due to current system implementations. This would require to both the back-end implementations and the interface controls to provide users with this capability. Currently the datasets integrated into the system are limited to a static set of their streaming counterparts however; the systems design supports dynamic data ranges. Testing would be required to determine how large of a data range the system could handle before performance hits would begin to degrade user experience. The required integration of the framework to support the collaborative and annotative capabilities into other systems is another limitation. Until other systems integrate the framework, users will be limited in their cross tool analysis. Users are also limited in the visualizations they can use to explore the datasets, as the system does not support dynamic dashboards or choice of visualization types. From these limitations, we can begin to identify areas of future work.

Future work for this thesis can be broken into two areas: 1) the models supporting the analytic processes and 2) the collaborative environment that supports it. Future work for models will focus on combining anomaly models and intervention models. Other work to be explored include the combination of sentiment analysis, frames, and clustering to define geo-political regions that share common framing strategies. Work in these areas will allow for better cueing and interesting event detection. Increased system intelligence can increase the capabilities of analysts and help improve the hypothesis generation process. By using the causality results to steer the statistical modeling and by applying more tests the accuracy of intervention and anomaly detection can be increased. Looking towards the collaborative environment, further improvements can be made to increase sense-making and common ground processes. The collaborative environment has been built with support for multiple visualization systems, however the application of this has not yet been completed. Future work for the collaborative environment includes the integration of other visualizations. A variety of visualizations can be integrated into the environment, which would allow for cross-tool analysis of data. Other work in this area is support for mobile devices and the capability to steer other visualizations using device specific controls. Extending support for mobile devices and device specific controls would allow users to collaborate in workspaces using any number of devices. These devices could then support partial views and controls of the visualizations as suitable to the devices resolution or scale.

REFERENCES

[1] P. André et al. "Continuum: Designing Timelines for Hierarchies, Relationships and Scale". In: *Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology*. UIST '07. Newport, Rhode Island, USA: ACM, 2007, pp. 101–110. DOI: 10.1145/1294211.1294229.

[2] *Armed Conflict Location & Event Data Project*. http://http://www.acleddata.com/. Accessed: 2015-03-28.

[3] *Armed Conflict Location & Event Data Project*. https://www.acleddata.com/wp-content/uploads/2015/01/ACLED_Codebook_2015.pdf. Accessed: 2015-03-28.

[4] S. Baccianella, A. Esuli, and F. Sebastiani. "SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining". In: *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*. Ed. by Nicoletta Calzolari (Conference Chair) et al. Valletta, Malta: European Language Resources Association (ELRA), May 2010.

[5] R. Bade, S. Schlechtweg, and S. Miksch. "Connecting Time-oriented Data and Information to a Coherent Interactive Visualization". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '04. Vienna, Austria: ACM, 2004, pp. 105–112. DOI: 10.1145/985692.985706.

[6] R. D. Benford and D. A. Snow. "Framing Processes and Social Movements: An Overview and Assessment". In: *Annual Review of Sociology* 26.1 (2000), pp. 611–639. DOI: 10.1146/annurev.soc.26.1.611.

[7] E. A. Bier, S. K. Card, and J. W. Bodnar. "Entity-based collaboration tools for intelligence analysis". In: *IEEE Symposium on Visual Analytics Science and Technology, 2008. VAST '08*. Oct. 2008, pp. 99–106. DOI: 10.1109/VAST.2008.4677362.

[8] M. Bogl et al. "Visual Analytics for Model Selection in Time Series Analysis". In: *IEEE Transactions on Visualization and Computer Graphics* 19.12 (Dec. 2013), pp. 2237–2246. DOI: 10.1109/TVCG.2013.222.

[9] J. Chae et al. "Spatiotemporal social media analytics for abnormal event detection and examination using seasonal-trend decomposition". In: *2012 IEEE Conference on Visual Analytics Science and Technology (VAST)*. Oct. 2012, pp. 143–152. DOI: 10.1109/VAST.2012.6400557.

[10] L. Chittaro and C. Combi. "Representation of temporal intervals and relations: Information visualization aspects and their evaluation". In: *In Proceedings of the Eigth International Symposium on Temporal Representation and Reasoning*. IEEE Computer Society Press, 2001, pp. 13–20.

[11] I. Cho et al. "VAiRoma: A Visual Analytics System for Making Sense of Places, Times, and Events in Roman History". In: *IEEE Transactions on Visualization and Computer Graphics* 22.1 (Jan. 2016), pp. 210–219.

[12] M. Cissel. "Media Framing: a comparative content analysis on mainstream and alternative news coverage of Occupy Wall Street." In: (2012), pp. 67–77.

[13] G. Convertino et al. "Supporting Content and Process Common Ground in Computer-supported Teamwork". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '09. Boston, MA, USA: ACM, 2009, pp. 2339–2348. DOI: 10.1145/1518701.1519059.

[14] W. Cui et al. "How Hierarchical Topics Evolve in Large Text Corpora". In: *IEEE Transactions on Visualization and Computer Graphics* 20.12 (Dec. 2014), pp. 2281–2290.

[15] W. Cui et al. "TextFlow: Towards Better Understanding of Evolving Topics in Text". In: *IEEE Transactions on Visualization and Computer Graphics* 17.12 (Dec. 2011), pp. 2412–2421.

[16] *Data Science Toolkit*. http://www.datasciencetoolkit.org. Accessed: 2015-03-18.

[17] N. A. Diakopoulos and D. A. Shamma. "Characterizing Debate Performance via Aggregated Twitter Sentiment". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '10. Atlanta, Georgia, USA: ACM, 2010, pp. 1195–1198. DOI: 10.1145/1753326.1753504.

[18] N. Diakopoulos, M. Naaman, and F. Kivran-Swaine. "Diamonds in the rough: Social media visual analytics for journalistic inquiry". In: *IEEE Symposium on Visual Analytics Science and Technology (VAST), 2010*. Oct. 2010, pp. 115–122. DOI: 10.1109/VAST.2010.5652922.

[19] 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.

[20] N. Diakopoulos et al. "Compare Clouds: Visualizing Text Corpora to Compare Media Frames". In: *Proceedings of IUI Workshop on Visual Text Analytics*. 2015.

[21] N. Diakopoulos et al. “Identifying and Analyzing Moral Evaluation Frames in Climate Change Blog Discourse”. In: *Proceedings of International Conference on Weblogs and Social Media (ICWSM)*. 2014.

[22] F. X. Diebold. *Elements of Forecasting*. South-Western College Publishing, 2006.

[23] P. S. Dodds et al. “Temporal Patterns of Happiness and Information in a Global Social Network: Hedonometrics and Twitter”. In: *PLoS ONE* 6.12 (Dec. 2011), e26752. DOI: 10.1371/journal.pone.0026752.

[24] M. Dörk et al. “A Visual Backchannel for Large-Scale Events”. In: *IEEE Transactions on Visualization and Computer Graphics* 16.6 (Nov. 2010), pp. 1129–1138. DOI: 10.1109/TVCG.2010.129.

[25] M. Dörk et al. “VisGets: Coordinated Visualizations for Web-based Information Exploration and Discovery”. In: *IEEE Transactions on Visualization and Computer Graphics* 14.6 (Nov. 2008), pp. 1205–1212. DOI: 10.1109/TVCG.2008.175.

[26] W. Dou et al. “LeadLine: Interactive visual analysis of text data through event identification and exploration”. In: *2012 IEEE Conference on Visual Analytics Science and Technology (VAST)*. Oct. 2012, pp. 93–102.

[27] K. Field and J. O’Brien. “Cartoblography: Experiments in Using and Organising the Spatial Context of Micro-blogging”. In: *Transactions in GIS* 14 (2010), pp. 5–23. DOI: 10.1111/j.1467-9671.2010.01210.x.

[28] D. Fisher et al. “Narratives: A visualization to track narrative events as they develop”. In: *2008 IEEE Symposium on Visual Analytics Science and Technology*. Oct. 2008, pp. 115–122.

[29] T. Gao et al. “NewsViews: An Automated Pipeline for Creating Custom Geovisualizations for News”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI ’14. Toronto, Ontario, Canada: ACM, 2014, pp. 3005–3014. DOI: 10.1145/2556288.2557228.

[30] D. Gotz and H. Stavropoulos. “DecisionFlow: Visual Analytics for High-Dimensional Temporal Event Sequence Data”. In: *IEEE Transactions on Visualization and Computer Graphics* 20.12 (Dec. 2014), pp. 1783–1792.

[31] C. W. J. Granger. “Investigating Causal Relations by Econometric Models and Cross-spectral Methods”. English. In: *Econometrica* 37.3 (1969), pp. 424–438.

[32] M. L. Gregory et al. “User-directed Sentiment Analysis: Visualizing the Affective Content of Documents”. In: *Proceedings of the Workshop on Sentiment and Subjectivity in Text*. SST ’06. Sydney, Australia: Association for Computational Linguistics, 2006, pp. 23–30.

[33] J. Heer, F. B. Viégas, and M. Wattenberg. “Voyagers and Voyeurs: Supporting Asynchronous Collaborative Visualization”. In: *Commun. ACM* 52.1 (Jan. 2009), pp. 87–97. DOI: 10.1145/1435417.1435439.

[34] M.J. Henry et al. “MultiFacet: A Faceted Interface for Browsing Large Multimedia Collections”. In: *2013 IEEE International Symposium on Multimedia (ISM)*. Dec. 2013, pp. 347–350. DOI: 10.1109/ISM.2013.66.

[35] S. M. Hsiang, M. Burke, and E. Miguel. “Quantifying the Influence of Climate on Human Conflict”. In: *Science* 341.6151 (2013). DOI: 10.1126/science.1235367. eprint: http://science.sciencemag.org/content/341/6151/1235367.full.pdf. URL: http://science.sciencemag.org/content/341/6151/1235367.

[36] 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. Paris, France: ACM, 2013, pp. 2707–2716. DOI: 10.1145/2470654.2481374.

[37] M. Itoh et al. “Analysis and visualization of temporal changes in bloggers’ activities and interests”. In: *2014 IEEE Pacific Visualization Symposium* 0 (2012), pp. 57–64.

[38] J. Kim et al. “Factful: Engaging Taxpayers in the Public Discussion of a Government Budget”. In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. CHI ’15. Seoul, Republic of Korea: ACM, 2015, pp. 2843–2852. DOI: 10.1145/2702123.2702352.

[39] Y. Kim, J. Hullman, and M. Agrawala. “Generating Personalized Spatial Analogies for Distances and Areas”. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. CHI ’16. San Jose, California, USA: ACM, 2016, pp. 38–48. DOI: 10.1145/2858036.2858440.

[40] R. E. Kraut, D. Gergle, and S. R. Fussell. “The Use of Visual Information in Shared Visual Spaces: Informing the Development of Virtual Co-presence”. In: *Proceedings of the 2002 ACM Conference on Computer Supported Cooperative Work*. CSCW ’02. New Orleans, Louisiana, USA: ACM, 2002, pp. 31–40. DOI: 10.1145/587078.587084.

[41] K. Krippendorff. “Agreement and Information in the Reliability of Coding”. In: *Communication Methods and Measures* 5.2 (2011), pp. 93–112. DOI: 10.1080/19312458.2011.568376.

[42] R. Langner, T. Horak, and R. Dachselt. “VISTILES: Coordinating and Combining Co-located Mobile Devices for Visual Data Exploration”. In: *IEEE Transactions on Visualization and Computer Graphics* PP.99 (2017), pp. 1–1. DOI: 10.1109/TVCG.2017.2744019.

[43] J. Li et al. “Social Media: New Perspectives to Improve Remote Sensing for Emergency Response”. In: *Proceedings of the IEEE* 105.10 (Oct. 2017), pp. 1900–1912. DOI: 10.1109/JPROC.2017.2684460.

[44] Y. Lu, F. Wang, and R. Maciejewski. “Business Intelligence from Social Media: A Study from the VAST Box Office Challenge”. In: *IEEE Computer Graphics and Applications* 34.5 (Sept. 2014), pp. 58–69. DOI: 10.1109/MCG.2014.61.

[45] Y. Lu et al. “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.

[46] D. Luo et al. “EventRiver: Visually Exploring Text Collections with Temporal References”. In: *IEEE Transactions on Visualization and Computer Graphics* 18.1 (Jan. 2012), pp. 93–105.

[47] A.M. MacEachren et al. “SensePlace2: GeoTwitter analytics support for situational awareness”. In: *2011 IEEE Conference on Visual Analytics Science and Technology (VAST).* Oct. 2011, pp. 181–190. DOI: 10.1109/VAST.2011.6102456.

[48] N. Mahyar and M. Tory. “Supporting Communication and Coordination in Collaborative Sensemaking”. In: *IEEE Transactions on Visualization and Computer Graphics* 20.12 (Dec. 2014), pp. 1633–1642. DOI: 10.1109/TVCG.2014.2346573.

[49] A. Malik et al. “A correlative analysis process in a visual analytics environment”. In: *2012 IEEE Conference on Visual Analytics Science and Technology (VAST).* Oct. 2012, pp. 33–42. DOI: 10.1109/VAST.2012.6400491.

[50] A. Malik et al. “Proactive Spatiotemporal Resource Allocation and Predictive Visual Analytics for Community Policing and Law Enforcement”. In: *IEEE Transactions on Visualization and Computer Graphics* 20.12 (Dec. 2014), pp. 1863–1872.

[51] C. D. Manning et al. “The Stanford CoreNLP Natural Language Processing Toolkit”. In: *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*. 2014, pp. 55–60.

[52] L. Mitchell et al. “The Geography of Happiness: Connecting Twitter Sentiment and Expression, Demographics, and Objective Characteristics of Place”. In: *PLoS ONE* 8.5 (May 2013), e64417. DOI: 10.1371/journal.pone.0064417.

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

[54] J. Neter et al. *Applied linear statistical models, 5th edition*. Vol. 4. Irwin Chicago, 1996.

[55] J. O'Loughlin, A. M. Linke, and F. D.W. Witmer. “Effects of temperature and precipitation variability on the risk of violence in sub-Saharan Africa, 1980–2012”. In: *Proceedings of the National Academy of Sciences* 111.47 (2014), pp. 16712–16717.

[56] Alexandra Olteanu et al. *Comparing Events Coverage in Online News and Social Media: The Case of Climate Change*. 2015.

[57] D. J. Peuquet et al. “A method for discovery and analysis of temporal patterns in complex event data”. In: *International Journal of Geographical Information Science* 29.9 (2015), pp. 1588–1611. DOI: 10.1080/13658816.2015.1042380.

[58] C. Plaisant et al. “LifeLines: Using Visualization to Enhance Navigation and Analysis of Patient Records”. In: *In Proceedings of the 1998 American Medical Informatic Association Annual Fall Symposium*. 1998, pp. 76–80.

[59] Jonathan C. Roberts. *Multiple view and multiform visualization*. 2000. DOI: 10.1117/12.378894.

[60] M. Sips et al. “A Visual Analytics Approach to Multiscale Exploration of Environmental Time Series”. In: *IEEE Transactions on Visualization and Computer Graphics* 18.12 (Dec. 2012), pp. 2899–2907. DOI: 10.1109/TVCG.2012.191.

[61] G. Sun et al. “EvoRiver: Visual Analysis of Topic Coopetition on Social Media”. In: *IEEE Transactions on Visualization and Computer Graphics* 20.12 (Dec. 2014), pp. 1753–1762. DOI: 10.1109/TVCG.2014.2346919.

[62] M. Thelwall et al. *Sentiment strength detection in short informal text*. 2010.

[63] D. Thom et al. “Spatiotemporal anomaly detection through visual analysis of geolocated Twitter messages”. In: *2012 IEEE Pacific Visualization Symposium (PacificVis)*. Feb. 2012, pp. 41–48. DOI: 10.1109/PacificVis.2012.6183572.

[64] B. Tomaszewski et al. “Supporting geographically-aware web document foraging and sensemaking”. In: *Computers, Environment and Urban Systems* 35.3 (2011), pp. 192–207.

[65] Feng Wang et al. “What’s In a Name? Data Linkage, Demography and Visual Analytics”. In: *EuroVis Workshop on Visual Analytics*. Ed. by M. Pohl and J. Roberts. The Eurographics Association, 2014. DOI: 10.2312/eurova.20141143.

[66] F. Wanner et al. “Visual sentiment analysis of rss news feeds featuring the us presidential election in 2008”. In: *In Workshop on Visual Interfaces to the Social and the Semantic Web (VISSW*. 2009.

[67] C. Weaver. “Cross-Filtered Views for Multidimensional Visual Analysis”. In: *IEEE Transactions on Visualization and Computer Graphics* 16.2 (Mar. 2010), pp. 192–204. DOI: 10.1109/TVCG.2009.94.

[68] C. Weaver. “Multidimensional visual analysis using cross-filtered views”. In: *IEEE Symposium on Visual Analytics Science and Technology, 2008. VAST ’08*. Oct. 2008, pp. 163–170. DOI: 10.1109/VAST.2008.4677370.

[69] B. Welch. “The generalization of ‘student’s’ problem when several different population variances are involved”. In: *Biometrika* (1947), pp. 28–35.

[70] W. Willett et al. “CommentSpace: Structured Support for Collaborative Visual Analysis”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI ’11. Vancouver, BC, Canada: ACM, 2011, pp. 3131–3140. DOI: 10.1145/1978942.1979407.

[71] Y. Woldemariam. “Sentiment analysis in a cross-media analysis framework”. In: *2016 IEEE International Conference on Big Data Analysis (ICBDA)*. Mar. 2016, pp. 1–5. DOI: 10.1109/ICBDA.2016.7509790.

[72] K. Wongsuphasawat and D. Gotz. “Exploring Flow, Factors, and Outcomes of Temporal Event Sequences with the Outflow Visualization”. In: *IEEE Transactions on Visualization and Computer Graphics* 18.12 (Dec. 2012), pp. 2659–2668. DOI: 10.1109/TVCG.2012.225.

[73] A. Wu et al. "Supporting Collaborative Sense-making in Emergency Management Through Geo-visualization". In: *Int. J. Hum.-Comput. Stud.* 71.1 (Jan. 2013), pp. 4–23. DOI: 10.1016/j.ijhcs.2012.07.007.

[74] Y. Wu et al. "OpinionFlow: Visual Analysis of Opinion Diffusion on Social Media". In: *IEEE Transactions on Visualization and Computer Graphics* 20.12 (Dec. 2014), pp. 1763–1772. DOI: 10.1109/TVCG.2014.2346920.