Who, When, Where and Why? Visualizing Civil Unrest Events

Parang Saraf*
Department of Computer Science
Virginia Tech

Nathan Self†
Department of Computer Science
Virginia Tech

Naren Ramakrishnan‡
Department of Computer Science
Virginia Tech

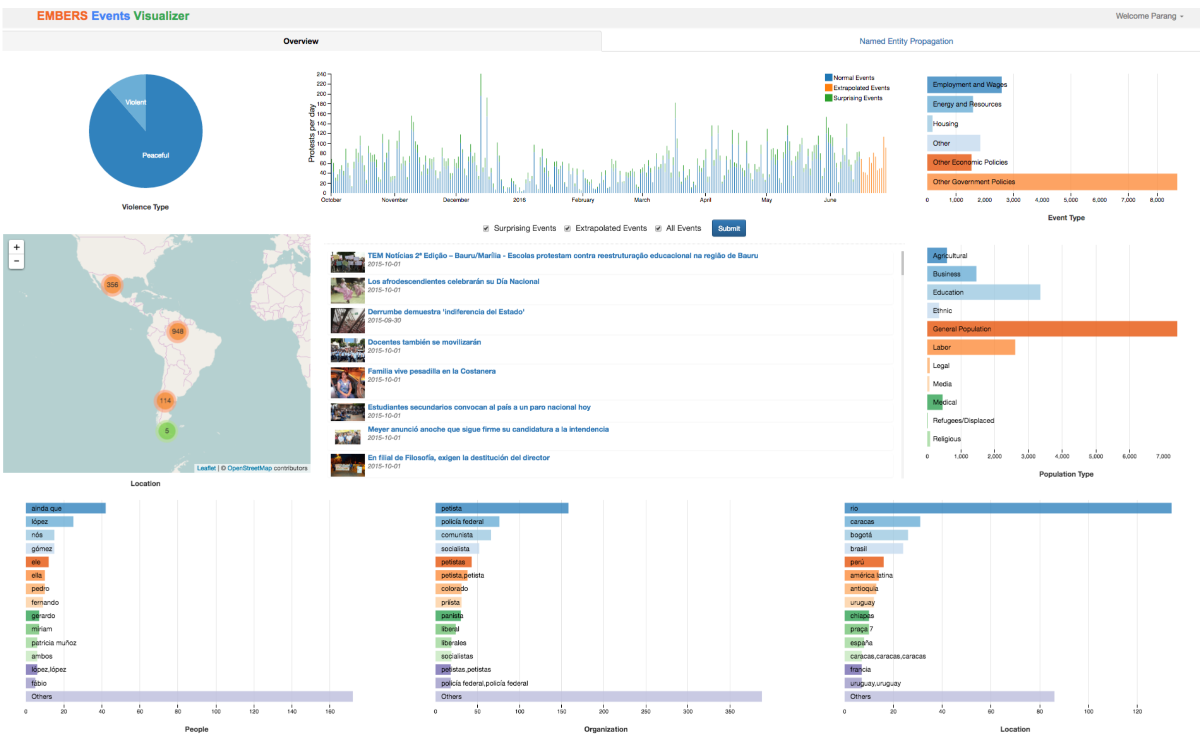

Fig. 1. **EMBERS Events Visualizer:** A framework with coordinated charts that allows analyst to club multiple query parameters in order to visualize spatio-temporal events data. The interface provides a timeline of civil unrest events and includes information about protest location, reason for protest, participating population group and the level of violence. It also displays news articles that reported these protest events along with named entities extracted from them. An analyst can apply successive filters on individual charts in order to analyze the progression of protest events. Finally, the analyst can also choose to examine surprising events or visualize extrapolated events.

Abstract— The paper describes a framework for analyzing spatio-temporal civil unrest events data of type who protested, when, where, and why. The framework employs coordinated charts to allow analysts to club multiple query parameters together. Additionally, an interface for visualizing propagation of named entities over time is also provided that enables an analyst to explore the evolution of associations between different entities related to a protest event. This framework allows analysts to examine civil unrest events at both macro and micro level.

Index Terms—Civil unrest events, spatio-temporal data, coordinated display

1 Introduction

News articles provide summaries of current events and include information such as what happened, when, where, and why. Since both spatial and temporal information is provided, events can be analyzed in a wide variety of ways and can enable both predictive as well as retrospective analysis. A database of such events allows scientific analysis of conventional wisdom by providing statistical evidences. For example, Washington Post has been working on curating a database of every fatal shooting in the United States by a police officer in the line of duty since Jan 1, 2015[1].

Many of the current information extraction systems make it feasible to extract structured information from news articles and generate a database of events. ICEWS [6], GDELT [4], and The Open Event Data Alliance (OEDA) [10] extracts political events from news articles. These extracted events are coded using the CAMEO [9] guidelines. Healthmap [2] extracts reports of global disease outbreaks from

*e-mail: parang@cs.vt.edu
†e-mail: nwself@cs.vt.edu
‡e-mail: naren@cs.vt.edu

Proceedings of the IEEE VIS 2016 Workshop on Temporal Sequential Event Analysis. Available online at: http://eventevent.github.io

[1]https://github.com/washingtonpost/data-police-shootings

online articles. RavenPack extracts financial events from news articles[2]. Saraf et al. [8] describes a semi-automated system for comprehensively extracting civil unrest events from news articles while minimizing the manual effort by 70%.

Once such events databases are created, the next logical step is to design systems that can analyze these extracted events. In this paper, we describe EMBERS Events Visualizer that allows visualization of civil unrest events of type 'who protested when, where and why?'. The system allows analysts to perform both macro and micro analyses of protest events. These protest events are generated by mining news articles in Spanish and Portuguese languages for 10 Latin American countries using the system described in [8]. An example of one such extracted event in shown in figure 2.

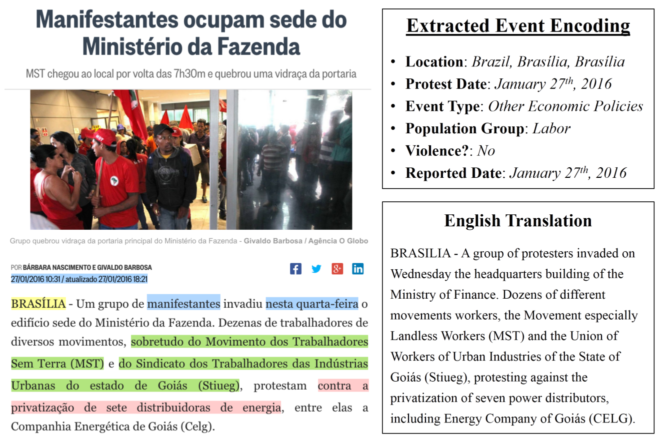

Fig. 2. A sample civil unrest event extracted from a portuguese news article from Brazil. The extracted event contains information about protest location, protest date, reason of protest, participating population group, and the level of violence. Additionally the article publish date is also identified.

Embers Events Visualizer provides following two views – 1) Coordinated View and 2) Named Entity Propagation View. Using these two interfaces, an analyst can answer queries like: How many protests happen on average per day? What are the most volatile time periods, and/or locations for protests? Which are the most active organizations participating in protests for a given country? What are the main reasons behind protests? Were the protests nationwide or localized? How do protests evolve over time? Does the involvement of a particular person or organization increases/decreases the chances of protest? For student led protests against government policies, who are the main leaders? How often do they protest? Additionally, the interface also provides mechanism for analyzing just the surprising events. Surprising events are identified using a maximum entropy approach. The coordinated view interface also allows analysts to view extrapolated events, which are generated based on the frequency of events in the past.

The rest of the paper is organized as follows: we begin by describing and comparing some of the widely used systems for visualizing spatio-temporal data in the the related work section. Thereafter we provide an overview of our system and describe individual features and components in section 3. This is followed by a use case that analyzes wide spread protest events in Brazil related to the impeachment process of Dilma Rousseff. Finally, we conclude our paper by describing some of the challenges and planned future work.

2 Related Work

In this section, we describe and compare some of the most prominent systems for visualizing spatio-temporal data. Spatio-temporal data includes information about location and time along with other attributes. GDELT [4] is one such system that visualizes spatio-temporal data related to political events. The data used by GDELT includes information of type *who did what to whom, when, and where?* and is very similar to the type of data used by our system. GDELT provides individual interfaces for visualizing different facets of data[3]. For example: *Event Timeline* interface shows event intensity by day, *Event Heatmap* interface visualizes spatial patterns, *Event TimeMapper* interface visualizes changes over time and space, and so on. However, all of these interfaces are stand alone and doesn't allow coordinated analysis of events.

European Media Monitor News Explorer [11] provides an interface for exploring current news[4]. It automatically determines what is being reported in the news, location of events, people involved and what they talked about. It allows filtering and clustering of news articles by location, people and topics. However, it is primarily an interface for exploring news articles and doesn't provide much analytical features for exploring events reported in news articles. Even though the interface has inter-connected components, it still doesn't allow clubbing of query parameters.

The Terrorism and Extreme Violence in United States (TEVUS) database and portal [3] provides an interactive, coordinated interface for visualizing and querying terrorist events based on the data collected from four open-source database. Even though the interface provides coordinated querying of data, an analyst can't apply multiple query parameters.

Recorded Future Cyber[5] system provides interconnected coordinated charts that allows user to apply multiple query parameters. However, the interface doesn't allow coordinated temporal or spatial analysis of events.

Another interesting visualization of spatio-temporal events is provided by Healthmap [2]. Healthmap provides an interface for visualizing spread and outbreak of diseases over time and location[6]. The system geocodes disease related news articles and also provides timelines for individual diseases. However, the charts are not inter-connected, thereby making it difficult to apply multiple filtering criteria.

As evident, most of the existing interfaces for visualizing spatio-temporal data focus primarily on presenting the data and provide minimal functionality in terms of performing any analytical exploration. The goal of EMBERS Events Visualizer is to not only present civil unrest data in an intuitive manner but also to provide functionality that can allow an analyst to apply various concurrent filtering criteria in order to understand and develop reasoning behind protest events.

3 System Overview

The system is comprised of two interfaces:

1. **Coordinated View:** As the name suggests, the coordinated view provides a series of inter-linked charts that allow an analyst to club multiple query parameters in succession. Applying a filtering criteria on one of the charts, updates all the remaining charts.
2. **Named Entity Propagation View:** This view allows an analyst to visualize the propagation of selected named entities and their associations over time.

We begin by providing a quick overview of the civil unrest dataset, followed by the description of the two interfaces.

3.1 Civil Unrest Dataset

An individual civil unrest event is characterized by following information – protest location, date of protest, reason for protest, participating population group(s) and whether the protest turned violent. Protest location is identified at city level. The reason for protest is classified into six categories – Employment and Wages, Energy and Resources, Housing, Other Economic Policies, Other Government policies and Other. Similarly, the participating population group is classified into eleven categories, namely Agriculture, Business, Education, Ethnic,

[2] http://www.ravenpack.com/products/

[3] http://analysis.gdeltproject.org/

[4] http://emm.newsexplorer.eu/NewsExplorer/home/en/latest.html

[5] https://go.recordedfuture.com/hubfs/data-sheets/cyber.pdf

[6] http://www.healthmap.org/en/

Labor, Legal, Media, Medical, Refugees, Religious and General Population.

Each of the civil unrest events has an associated news article that reported the event. These events were extracted by mining Spanish and Portuguese news articles from 10 Latin American countries beginning October 2015 till present using the system described in [8]. The 10 Latin American countries for which events were extracted are as follows: Argentina, Brazil, Chile, Colombia, Ecuador, El Salvador, Mexico, Paraguay, Uruguay and Venezuela.

3.2 Coordinated View

The interface provides inter-connected coordinated charts for displaying each of the individual components of a protest event along with the associated news article. These news articles are passed through a natural language processing toolkit[7] in order to extract named entities from them. We focus primarily on people, organization and location. The interface also displays these named entities in inter-connected charts in order to help an analyst understand the key players associated with the protest.

All the charts are dynamic and interactive in nature. Since, the charts are inter-connected, applying a filtering criteria on one of the charts updates all the remaining components of the interface. This greatly helps in applying successive filtering criteria in order to drill down to particular events. The protest timeline chart provides an easy to use intuitive 'brush' functionality in order to filter protest events by time period.

The geographical map shows events at city level and clusters events together based on the level of zoom. As an analyst zooms in, the map gets updated to show the geographical spread of events. An analyst can also click on certain countries on the map to show protest events for only those countries.

The table listing news articles gets updated based on the filtering criteria. There are two views for news articles – summary and detailed view. The summary view lists all the news articles matching the filtering criteria. Each news article is accompanied by its associated image, news source and date of publication. Clicking any news article opens up the detailed view of the article in a pop-up modal. The modal shows the associated article image along with the full text and link to the original article.

In addition to visualizing civil unrest events, the system also provides two additional features – Surprising Events and Extrapolated Events.

3.2.1 Surprising Events

The civil unrest dataset contains a mix of everyday mundane sporadic events as well as surprising events such as Brazilian Spring that are widespread, spans several days and are organized at unprecedented scale. Needless to say, identifying and analyzing such events are of vital importance to analysts.

In order to identify surprising events, we employ a maximum entropy based approach as described in [5]. In this approach, for each event we use location (at country level), reason of protest and protesting population group. The civil unrest dataset can then be visualized as a cube of counts as shown in figure 3. A maximum entropy distribution conditioned on the marginals of the cube is inferred by using iterative proportional fitting [1]. Surprising events are identified for each month. In order to do so, events from the last three months are used to populate the cube and then iterative proportional fitting is used to generate maxent values of each cell of the cube. These estimated values are then scaled such that the total count matches the total event count of the given month. Finally the surprising cells are identified by comparing the counts in each cell for a given month with the estimated maxent counts. Significant cells with difference greater than five standard deviation are labeled as surprising cells and the events corresponding to these cells are marked surprising events.

Please note that surprising events are events from the civil unrest dataset that have been specifically identified using statistical measures.

[7]https://www.rosette.com/

When 'all events' is selected on the interface all the events including surprising events are shown in the same color. However, when 'surprising events' is also selected in addition to 'all events', then surprising events are filtered out from all events and shown in a different color on the protest timeline.

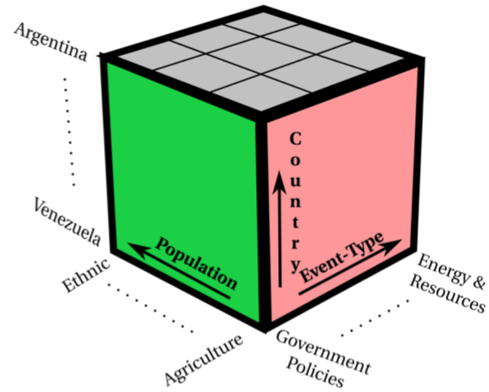

Fig. 3. **Identifying surprising events:** for a given month, a maximum entropy cube is generated by using location, reason and population from the events that happened in the last three months. Maxent values for individual cells in this cube are estimated using iterative proportional filtering. Surprising events are determined by identifying cells having difference in scaled estimated values and true values greater than five standard deviation.

3.2.2 Extrapolated Events

An effective complementary feature for an interface that visualizes historical events, is the ability to visualize extrapolated events, so that predictive trends and patterns be identified. With this aim, the interface also provides a functionality to visualize extrapolated events. The events are extrapolated using the historical event frequencies for a particular country as explained in [7]. A random uniform distribution function is used to assign dates to the events. Since the events are extrapolated there are no corresponding news articles and hence in case if the analyst chooses to visualize only the extrapolated events, then the plots for named entities as well as table of news articles are left blank.

3.3 Named Entity Propagation View

This interface (fig. 4) primarily aims at exploring progression of news articles that are clustered by named entities. An analyst begins by selecting a country, time period, number of days for each cluster and an initial set of named entities to track. Based on the selected criteria, the system filters out news articles and clusters them into groups based on the selected number of days. For each of these individual clusters, the system then displays named entity word clouds. The interface provides two modes for visualizing the propagation – automatic and manual. In the automatic mode, the named entity word clouds get propagated every three seconds while in the manual mode, the analyst can click next or back to control the propagation.

An interface like this helps an analyst uncover the associations between people, organizations and locations and understand how this association has evolved over time. The analyst can also click a particular news article to see the detailed view.

4 Use Case

For our use case, we will focus on 2015-16 protests that took place in Brazil that led to the initiation of impeachment process against their president Ms. Dilma Rousseff[8,9]. The protests began in March 2015 when it was discovered that several ministers of the ruling party were involved in corruption and bribery and concluded in May 2016 with the suspension of Ms. Rousseff. We will specifically examine

[8]https://en.wikipedia.org/wiki/Impeachment_process_against_Dilma_Rousseff
[9]https://en.wikipedia.org/wiki/2015–2016_protests_in_Brazil

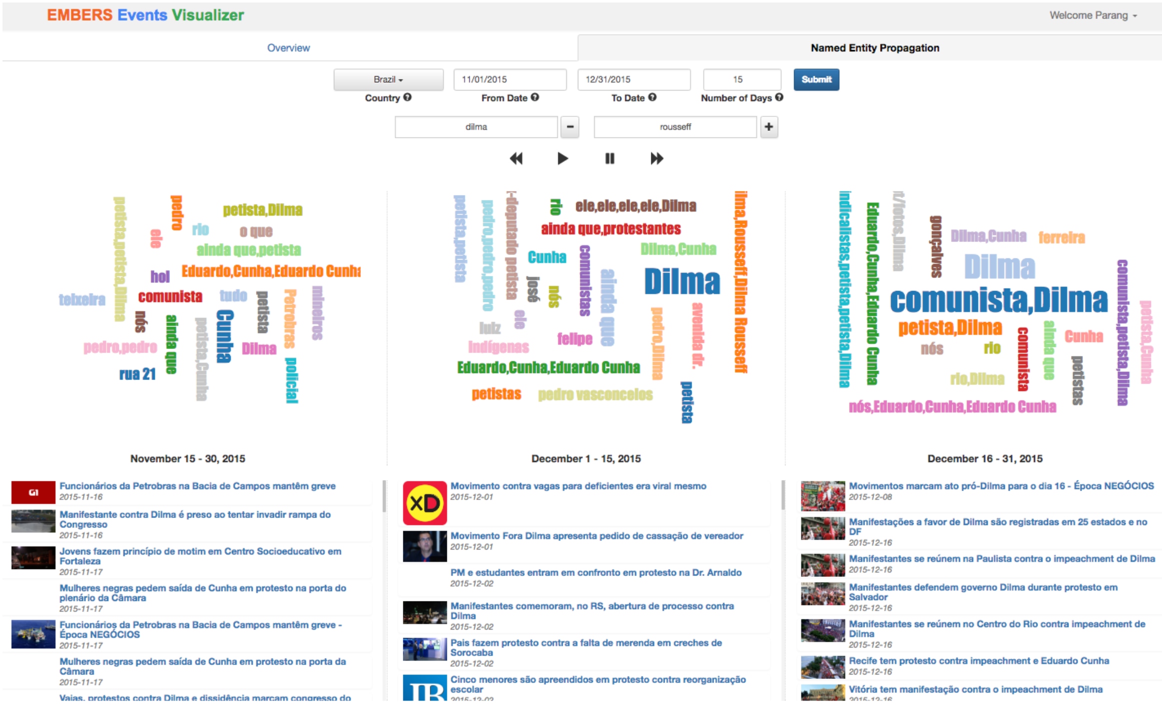

Fig. 4. **Named Entity Propagation View:** This interface aims at visualizing progression of news articles that are clustered by named entities. The progration of named entities can be viewed in either automatic or manual model. This view helps an analyst uncover related entities and makes it easy to understand how the association between named entities has evolved over time

the events during her impeachment process that began on December 2^{nd}, 2015.

We begin by applying a filter on the geographical map to show only the protest events in Brazil. The spatial spread of events tells us that São Paulo and Rio de Janerio are the hotspots for protests. The timeline chart along with surprising events shows that there is an unusual uptick in protests beginning November 15. So, we apply second filter on the timeline chart to restrict the protest events from November 1 – Jan 15, 2016. Doing so updates all the remaining charts to reflect the reason of protest, protesting population groups and the key named entities. Next we analyze the same event on the named entities propagation view. We restrict ourselves to Nov 1 – Dec 31, 2015 time period for Brazil, filter articles by 'Dilma' and 'Rousseff' and use a 15 day clustering window. Looking at the individual clusters for these time periods helps an analyst understand how named entities appear and disappear over time.

5 Conclusion

With an increase in information extraction systems that can extract socio-political events from news articles, it is essential to explore visualizations that can work on the extracted events and simplify the analytical and inference generation processes for an analyst. Embers Events Visualizer is one such system that employs coordinated charts in order to allow an analyst to use multiple query parameters in succession. With Embers Events Visualizer, analysts can uncover patterns and associations between named entities associated with civil unrest events in an intuitive manner. An interesting supplementary feature for this system will be an interface that can make it easy to chain related news stories together in order to build a timeline of events, specially for the ones that span across multiple days/months.

Acknowledgments

This work was supported by the Intelligence Advanced Research Projects Activity (IARPA) via DoI/NBC contract number D12PC000337. The US Government is authorized to reproduce and distribute reprints of this work for Governmental purposes notwithstanding any copyright annotation thereon. Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DoI/NBC, or the US Government.

References

[1] Y. M. Bishop, S. E. Fienberg, and P. W. Holland. *Discrete multivariate analysis: theory and practice*. Springer Science & Business Media, 2007.

[2] C. C. Freifeld, K. D. Mandl, B. Y. Reis, and J. S. Brownstein. Healthmap: global infectious disease monitoring through automated classification and visualization of internet media reports. *Journal of the American Medical Informatics Association*, 15(2):150–157, 2008.

[3] G. LaFree and W. Braniff. Project fact sheet: Terrorism and extremist violence in the u.s. database and portal. 2014.

[4] K. Leetaru and P. A. Schrodt. Gdelt: Global data on events, location, and tone, 1979–2012. In *ISA Annual Convention*, volume 2. Citeseer, 2013.

[5] S. Muthiah, P. Butler, R. P. Khandpur, P. Saraf, and N. Self et al. Embers at 4 years: Experiences operating an open source indicators forecasting system. In *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pages 205–214, New York, NY, USA, 2016. ACM.

[6] S. P. Obrien. Crisis early warning and decision support: Contemporary approaches and thoughts on future research. *International Studies Review*, 12(1):87–104, 2010.

[7] N. Ramakrishnan, P. Butler, S. Muthiah, N. Self, and R. Khandpur et al. 'beating the news' with embers: Forecasting civil unrest using open source indicators. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '14, pages 1799–1808, New York, NY, USA, 2014. ACM.

[8] P. Saraf and N. Ramakrishnan. Embers autogsr: Automated coding of civil unrest events. In *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pages 599–608, New York, NY, USA, 2016. ACM.

[9] P. A. Schrodt. Cameo: Conflict and mediation event observations event and actor codebook. *Pennsylvania State University*, 2012.

[10] P. A. Schrodt, J. Beieler, and M. Idris. Threesa charm?: Open event data coding with el: Diablo, petrarch, and the open event data alliance. In *ISA Annual Convention*, 2014.

[11] R. Steinberger, B. Pouliquen, and E. van der Goot. An introduction to the europe media monitor family of applications. In *Proceedings of the 2009 SIGIR Workshop*, SIGIR-CLIR, pages 1–8, July 2009.