Peña-Araya et al. *EPJ Data Science* (2017) 6:25
DOI 10.1140/epjds/s13688-017-0122-8

EPJ Data Science
a SpringerOpen Journal

EPJ.org

REGULAR ARTICLE Open Access

Gaining historical and international relations insights from social media: spatio-temporal real-world news analysis using Twitter

Vanessa Peña-Araya[1*], Mauricio Quezada[1], Barbara Poblete[1] and Denis Parra[2]

*Correspondence:
vpena@dcc.uchile.cl
[1]Department of Computer Science, University of Chile, Avenida Beauchef 851, Santiago, Chile
Full list of author information is available at the end of the article

Abstract

The immense growth of the social Web, which has made a large amount of user data easily and publicly available, has opened a whole new spectrum for research in social behavioral sciences. However, as the volume of social media content increases at a very fast rate, it becomes extremely difficult to systematically obtain high-level information from this data. As a consequence, tasks related to the analysis of historical news events based on social media data have not been explored, which limits any type of comparative historical research, causality analysis, and discovery of knowledge from patterns extracted from aggregated social media event information.

In this work, we target this issue by proposing a compact high-level representation of news events using social media information. This representation explicitly includes temporal information about the event and information about locations, in particular of geopolitical entities. We call this a *spatio-temporal context-aware event representation*. Our hypothesis is that by including social, temporal, and spatial information in the event representation, we are enabling the analysis of historical world news from a social and geopolitical perspective. This facilitates, new information retrieval tasks related to historical event information extraction and international relations analysis. We support our claims by presenting two applications of this idea: the first, a visual tool, named Galean, for retrieval and exploration of historical news events within their geopolitical and temporal context. The second, a quantitative analysis of a 2-year Twitter dataset of news events reported by U.S. and U.K. media, which we explore using data mining techniques on our event representations. We present two case studies of event exploration using Galean and user evaluation of this tool, as well as details of our data mining empirical results.

Keywords: visualization; geo-temporal context; event modeling; historical analysis

1 Introduction

As online social networks become massively popular, they are used as reliable and efficient news sources. Many users exploit social media platforms to obtain information, especially breaking news [1]. Even traditional mass media organizations such as newspapers and TV news channels now use social media platforms to inform their audience more quickly. Social media users are not only consumers of this information, but also producers and broad-

casters. Millions of people from all over the world have assumed the task of reporting and commenting on newsworthy events. In particular, the social platform Twitter [2] has become a preferred source for users to find up-to-date information. The messages published in Twitter are called *tweets* and are constrained to 140-characters. When breaking news occurs, Twitter users quickly react by generating content and producing interactions. The particular nature of Twitter messages, as well as the fact that most of its users use the platform from mobile devices, facilitates extremely fast information propagation.

Twitter provides excellent conditions for social behavior analysis, as well as comparative historical research, among many other social and scientific disciplines. In particular, the field of *comparative historical research* examines historical events in comparison to other historical events to gain general knowledge that goes beyond a particular event [3]. So far, historical research had been restricted to traditional archival data and historians' written account of past events. Nevertheless, it is undeniable that the data poured into social media about world events is of great value to society. The proof is in the increasing body of scientific work surrounding retrospective microblog data. Just to mention a few examples: Castillo et al. [4] extracted information to predict the credibility of rumors in social media, Sakaki et al. [5] used Twitter for real-time earthquake detection, Pak and Paroubek [6] studied Twitter messages as a corpus for sentiment analysis, and Saravanou et al. [7] used tweet coordinates to find locations that have been affected by floods.

Despite the usefulness of historical information extracted from social media, there is not much research addressing the topic of retrospective analysis of this data. Social media in general, and Twitter in particular, produce streaming data that is volatile, which most likely explains why existing research concentrates only on a particular event, such as an earthquake or on predefined datasets.

We addressed this issue by introducing a visual tool and novel data mining tasks, based on a compact representation of real-world news events. This representation was designed to summarize information about real-world events from social media data, which is enhanced with the event's geographical and temporal context.

Our event representation incorporates two types of spatial data about an event: (1) locations directly *involved* in the real-world event occurrence (i.e., the main places that are mentioned in messages about the event), which we refer to as *protagonist locations*, and (2) locations from where social network users *comment* on the event (i.e., the places where users that comment are located), which we refer to as *interested locations*. For example, when an earthquake took place in Nepal in April, 2015, most of the messages mentioned Nepal, which indicated that this was the location where the event had taken place. Therefore, if we consider locations at a country level, Nepal can be regarded as the *protagonist location* of that event. However, the users that posted the messages about Nepal were distributed all over the world, indicating that this event had global impact. Furthermore, some countries had more users interested in the event than other countries, such as, neighboring countries and countries with citizens among the victims. These would be considered as the *interested locations* of that event.

Our work is based on the hypothesis that by adding spatio-temporal context to news events, such as protagonist and interested locations, and the time at which it occurred, we can discover new information based solely on social media data. In particular, the application of our event representation allows us to find relationships among events and among locations, such as:

(i) **event similarity:**
- **based on their protagonist locations**, i.e., retrieve all the events that occurred in certain location, or that directly involved similar groups of locations;
- **based on the locations that are interested in the event**, i.e., retrieve all of the events that produced the similar interest in other locations.

(ii) **location similarity:**
- **based on events in which a location is protagonist**, i.e., retrieve the locations that are protagonists in the same events;
- **based on their interest in events**, i.e., retrieve sets of locations that showed similar levels of interest in the same events.

(iii) **any combination of the above.**

These similarity relationships along with temporal context can facilitate the implementation of novel information retrieval tasks. These tasks include: event search, event understanding, geopolitical analysis, international relations analysis (when considering locations at a country level), historical comparative analysis, among others.

Our contribution:

1 We introduce a visual tool, named Galean, for exploration of historical news event collections based on our proposed approach. This tool allows the user to view event evolution over long periods of time, the relationship between the geopolitical entities that participate in them, and emergent patterns;
2 We present a novel high-level representation of news events based on information extracted from social media. This representation emphasizes the geographical and temporal context of news.
3 We present an exploratory analysis of a 2-year data collection in which we use our proposed event representation to identify connections and similarity patterns among countries.

In this article, we describe our news event representation, our visual exploration tool and an exploratory analysis of a real-world collection. We present a user evaluation that provides evidence of the usefulness of Galean for information retrieval tasks related to spatio-temporal event exploration. We present two case studies of real-world news events that illustrate the use of our tool. We show how our visualization can facilitate manual event or location (geopolitical entity) tracking over time, based on social media. In addition, we provide an empirical analysis of a real-world news event collection extracted from Twitter. We discuss our main findings regarding the international relations that can be obtained from the complete dataset. These relations show well known relationships among countries, as well as new information that reflects how the Twitter community is affected by different events.

The rest of the paper is organized as follows. Section 2 describes relevant research related to our approach. Section 3 presents our proposed spatio-temporal context-aware news event representation. Section 4 describes the visualization tool and application framework. Section 6 describes our exploratory data mining analysis using the proposed high-level event representation. Section 7 discusses our findings and known limitations of our work. Finally, Section 8 presents conclusions and future work.

Relation to previously published work. This paper is an extended version of our preliminary work on spatio-temporal context-aware event representation and visualization [8, 9]. Specifically, Section 3 extends Quezada et al. [8] by providing a more formal and

detailed description of event representations. Section 4 is an extension of the visualization introduced as a demonstration paper in Pena-Araya et al. [9], which we expand by adding much more detail on our system design, two new case studies, as well as a quantitative evaluation. Section 6 provides a completely new data mining analysis that uses the proposed event abstraction. In addition, we also mention our earlier work, Maldonado et al. [10] and Kalyanam et al. [11], which study news events as well, but we do not extend those works in this current article. The remaining sections of this paper are entirely new.

2 Related work

In this section, we discuss relevant prior work for our main topics of research, which are: (1) modeling real-world events using the Web and/or social media, with spatio-temporal context information, (2) quantitative historical event analysis, and (3) visualization of news and events within their geographical and temporal scope. We note that, although our work also involves event detection, collection of event data, and geolocating this information, our current contribution is not centered in those areas. Therefore, the literature related to those topics is discussed in other sections as required to understand specific details.

2.1 Event models using social media

Most of the research in social media event analysis has been directed towards the creation of event models for specific tasks such as detection, tracking, summarization and characterization of events in social media streams. However, not much work has been focused on high-level event modeling with context information, such as spatio-temporal information.

In the work of Kamath et al. [12], Twitter *hashtags* (i.e., user-generated string prefixed by # that users add to tweets as a way to associate it with an event or a topic) were analyzed in a large-scale study of the spatio-temporal dynamics of *memes*. In this work a hashtag was represented as a tuple consisting of the coordinates of the hashtag's location over time. They used a simple model to find interesting insights about the adoption and spread of memes in social media. Memes are information which emerges from social networks and spreads in a viral way. However, meme dissemination does not necessarily resemble how other types of information will propagate, such as information about events that originate outside of the network (i.e., exogenous events). Following this motivation, Kalyanam et al. [11] studied how exogenous events, in this case real-world news, propagate in social media. In their work, they modeled news events based on the interarrival time between social media posts, without considering any of the geographical information associated to the event. Their goal was to model the intensity of the user activity that is triggered by a real-world news event. Though, in our current work we also study real-world news using the same data extraction technique as Kalyanam et al. [11], our approach differs in that our event model is not based on the interarrival times of tweets, but rather on the geographical context of social media information.

In a different type of study, Leetaru [13] performed a large-scale analysis of 30 years of digitized news articles. The author computed sentiment scores and geolocation for each article. The study indicated that some critical events in the past, such as social revolutions, could have been forecasted by looking at sentiment scores over time. In addition, the author performed community detection on country graphs by analyzing news in which two or more countries were involved. In this sense, our approach is similar, because we model countries in terms of their co-occurrence in news. However, our work is focused on automatic information extraction from online social streams and on the creation of a more

general representation. We do not focus on the analysis of sentiment of edited content from formal news media outlets, but on the interactions between locations, based on the aggregated reactions and opinions of users of social platforms.

There are other approaches for event information modeling, which come from the area of automatic text summarization. Chakrabarti and Punera [14], for example, used hidden Markov models to represent sub-events, within a broader event that is described using Twitter data. This model identified sub-events based on the burstiness of the input data stream and the word distribution of the main event. Another approach was presented by Quezada and Poblete [15], which focused on automatic summarization of multimedia content by using social media posts as surrogate text for multimedia documents. A similar approach was used by Alonso et al. [16], which was based on the *social signature* of documents (that is, the set of keywords of social media messages that point to a document), to augment the document information.

Several other features that have been used for modeling events on social media are worth mentioning, such as, users involved in an event [17], the credibility of the information that is published [4], and latent sentiment of content [18], among others. In addition, temporal features for events have been used in tasks such as, the detection of events based on the temporal dynamics of their mentions in social media [19], and also for event categorization [20]. Nevertheless, we do not use those features at this time.

Certain studies focused specifically on the task of detecting events and tagging their relevant geolocations. In particular, some works targeted the detection of localized events [21–25], others the detection of global events [26], and the detection of critical events [5, 27]. Dong et al. [28], specifically, considered that events had different temporal and spatial scales and proposed a multi-scale event detection approach for social media. This approach focuses on detecting and reporting events with geolocalization. Our current approach differs from existing work, in that we create an aggregated representation of the information about real-world events, producing a high-level representation that includes the event's geographical context, which is extracted from social media. In addition, we enrich the information about an event by using the locations of the users that post information about it.

Wang et al. [17] visualized topics based on the extraction of geographical entities from tweet text. They did not use this information to establish the location of an event, but rather for event exploration. SensePlace2 [29] is a Visual Analytics tool that allows users to explore a set of tweets and models them by showing two geographical types of information: the locations from where users discussed the topic and the locations being mentioned in tweets. However, unlike our work, this information was only used at single tweet level, and not at event level.

In the domain of cyber-physical systems, *events* are viewed as conditions of interest [30] within a cyber-physical system, or as the co-occurrence of two people in the same physical place [31]. In general, events are modeled according to the state of the objects in the system, considering attributes, time and location. The work presented by Tan et al. [30] bears certain similarities with our own, in the sense that they considered an event to encompasses multiple information about a condition of interest in the system (in our case in the online social network), including time and physical locations. In addition, the authors defined different kinds of temporal and geographical scopes for their events, which are similar to our definition of *event impact*. The main difference relies in that our approach aims to

capture high-level information of how a complex exogenous event, such as a news event, is perceived by social network users in an aggregated way. Therefore, we focus on geopolitical divisions as units of aggregated spatial information and on representing geopolitical interactions.

Despite that the idea of adding spatio-temporal context to social media data is not novel, to the best of our knowledge our work is the first that formally introduces *protagonist* and *interested* locations in a high-level event representation. The novelty of our approach relies on the extension of the notion of spatial context, first by associating real-world news to one or more protagonist locations, and second by associating real-world news to the locations where they generated interest. In addition, our work does not focus on event detection, classification or summarization, as most of the prior work on event analysis does.

2.2 Quantitative historical event analysis

We provide a revision of the literature on *quantitative history* research applied to event analysis and social media. Quantitative history is an approach to historical research that makes use of quantitative and digital tools [32]. To the best of our knowledge, our work is the first to make use of social media data for quantitative historical research.

Prior work used digitized newspapers and books for extracting quantitative knowledge [13, 33, 34]. Michel et al. [33] built a corpus of 5 million books and analyzed them using word frequencies to investigate cultural trends, and called this type of study "Culturomics". Leetaru [13] performed a large-scale study of 30 years of digitized newspapers, described in the previous section. Chadefaux [34] used a dataset from Google News Archive to predict military conflicts.

A different line of research covers digitized writings and the Semantic Web. Suchanek and Preda [35] proposed the study of "Semantic Culturomics", in which the analysis of newspapers should go beyond the study of word frequencies in order to integrate knowledge bases (such as DBPedia [36]) to answer complex user queries. Additional research has used knowledge bases along with human writings, such as newspapers [37, 38]. A survey on this topic is provided by Meroño-Peñuela et al. [39].

Compared to prior work, our approach is the first to consider user-generated information networks, such as online social networks, which are a growing data source at much larger scale. We consider that social media can provide additional and novel information to that found in news articles and books. User-generated content, reflects social opinions and points of view related to current world-events. This content is generated in real-time, it is not edited and does not depend on the editorial lines of formal news outlets. We believe that these unique characteristics make social media a challenging and valuable source of historical information. Our approach incorporates the content of social media platforms about real-world news, as well as aggregated geographical information that conveys the importance and scope of these events.

2.3 Geographical based visualizations of social network data

There are several visualization tools that show where a news event has happened or from where social media users are commenting on it. In this section, we review the tools that are relevant for our work, mainly focused on what type of geographical information they convey and what users can obtain from them.

If an event is represented as a set of documents, then one way of understanding this event is by using the documents metadata. There are several visualization tools based on

this idea, which show the geographical distribution of documents, allowing users to answer specific questions. Some examples are TwitInfo [40], Jasmine [21], and others [41–43]. Some systems provide filters for users to select documents published from particular places at particular times. For example, ScatterBlogs2 [44] is a visual analytics system for understanding messages from Twitter that allows users to interactively filter messages by their geographical and temporal context, using the coordinates from where the message was emitted. Also, Bosch et al. [45] created a system that aims to help users analyze social media using various sources, including search and filtering features for messages in their spatial and temporal dimensions. All of these systems use a map to display the geographical distribution of messages, (or of users) in order to describe a topic or event. Whisper [46] uses a different metaphor: by representing messages of an event as seeds of a sunflower, a user can follow how information disseminates by viewing the locations from where people commented on an event, or from where a message was re-posted. In contrast to these approaches, which are centered on user messages, our visual tool focuses on the characteristics of an event as a whole, providing details (messages) on demand.

There are also visualization systems for describing events. Visgets [47] provides a visual interface to represent entities from different data sources, such as the ACM WWW proceedings or the social news site Global Voices Online [48]. A user can search and filter entities by time, space and keywords. Visgets represents entities by their geographical location using entity metadata. LeadLine [17] is an interactive visual analytics system that supports the exploration of events detected automatically from news and social media. The LeadLine system extracts places mentioned in messages to identify where a piece of news was relevant. Event Registry [49] is a system that monitors media sources to detect news events in more than ten languages. It also presents a map visualization that displays each event as a bubble over the location where it happened. Our approach complements these systems by leveraging event information and the impact that its information had in social media.

SensePlace2 [29] is a web system that shows locations mentioned in tweets and the locations from where these tweets were published. From that point of view, SensePlace2 is a system that allows users to ask: *"what places are involved in an event and from where are people commenting on it?"*. Therefore, it is the most similar system to the work presented in this paper. However, as the authors of SensePlace2 described in their own work, the main limitation of their tool is that it focuses more on the dimensions of the events rather than on the events themselves. Our work complements SensePlace2 by: (1) focusing on overall event information, (2) by allowing users to explore relationships among countries, and (3) by showing the user if news events are local or international. Another system that is similar to ours is The News Co-occurrence Globe [50], which displays the co-occurrence of countries in news media reports on a 3D map. However, it does not currently provide the functionality to put focus on events. Our work allows the user to focus on events to see how relationships between countries are created and evolve over time.

In summary, to represent events in their geo-temporal context, most visualization systems either show the geographical distribution of the documents that discuss news events, or the information about the event itself. However, these approaches are limited if the user needs to retrieve news events or ask complex questions such as *where did event x happen?*, *how did people around the world react to event x in social media?*, *did event x impact only locally or did it have global impact?*, *which countries showed the most interest in event x?*

or *have other countries also been involved in similar events to x?* In particular, to the best of our knowledge, our is the first approach to consider that events can be related to multiple locations, reflecting interactions between geopolitical entities. Overall, our tool is the first to allow historical news exploration and retrieval that considers the temporal and spatial context of the user and of the event. In addition, providing the means for manual exploration of vast amounts of contextualized events described using social media data.

3 Event representation

We introduce a novel high-level event representation, called *spatio-temporal context-aware event representation*, with the purpose of gaining insight about real-world news from social media as well as from the relations between locations and impact that these events induce. Specifically, we define our event representation and how it can be used to study relations among locations.

3.1 Event representation definition

We represent an event as a complex information unit that encompasses all of the available social media content associated with a certain news topic, as well as its aggregated spatial and geographical information. In particular, we incorporate information about the locations involved in the event occurrence, and the locations of the users that post messages about the event. This representation is solely based on the social media activity surrounding the event in the online social network, without including any external information sources. Specifically, we define two types of spatial contexts, which we call:

1 **protagonist locations**, which are the locations involved in the event, and
2 **interested locations**, which are the locations from where users comment on the event.

For example, consider the news about Chile and Peru's maritime dispute at The Hague in The Netherlands [51]. If we define locations at country level, then this is an event for which Twitter users mention mostly three countries when discussing the event: Chile, Peru and The Netherlands (other mentions are negligible). Hence, according to our definition this event is considered to have three protagonist locations. However, users that comment on this event are located mostly in: Chile, Peru, Argentina and Bolivia. Therefore, the event is considered to have four interested locations.

More formally, we define an event E as a tuple of the form:

$$E = (K, D, T, \mathbf{P}, \mathbf{I}), \tag{1}$$

where K is a set of keywords, which succinctly describe the news topic, D is the date of the event detection, T is a set of tweets about the event, published by users of online social networks. In addition, consider $L = \{l_1, l_2, \ldots, l_{|L|}\}$ to be the set of existing locations. We augment the information about the event by explicitly including its spatial context with the vectors $\mathbf{P}$ and $\mathbf{I}$, which correspond to the *protagonist* and *interested* location values, respectively, for the event E. This is, the j-th dimension of vector $\mathbf{P}$ contains the number of times that the location l_j is mentioned by the tweets in T. On the other hand, the j-th dimension of vector $\mathbf{I}$ contains the number of tweets in T that were posted by users in the location l_j.

Using the information introduced by vectors $\mathbf{P}$ and $\mathbf{I}$ we can derive the scope of an event E from two perspectives, *provenance* and *impact*, defined as follows:

- **Provenance:** Indicates whether an event is local, regional or global in terms of the locations it involves. We consider an event to be of *local provenance* if it involves only one protagonist location. We consider an event to be of *regional provenance* if it involves two or more protagonist locations that are all from a same region (e.g., for countries, this means neighboring countries or from a same continent[a]). We consider an event to be of *global provenance* if it involves two or more protagonist locations in which at least one is not from the same region. Vector **P** contains this information for a given event E.
- **Impact:** Indicates if an event is local, regional or global in terms of how many locations show interest in it. We consider an event to be of *local impact* if it generates conversation from users in only one location (i.e., one interested location). We consider an event to be of *regional impact* if it generates conversation from users in more than one interested location, all from the same region. We consider an event to be of *global impact* if it generates conversation from users in more than one interested location in which at least one of those locations is not from the same region. Vector **I** contains this information for a given event E.

For example, the message *"Australia confirms signals detected by China 'consistent' w/ #MH370 black box"*, discussed an event in which Australia and China are involved. Therefore, Australia and China can be considered as protagonist locations in this event. On the other hand, this particular news event was discussed extensively by users located in several countries, including: USA, Canada, Colombia, U.K., India, Nigeria, South Africa, Indonesia, Australia, France, Germany, China and Italy. Therefore, this is an event that had *global provenance* (i.e., more than one protagonist location from different regions) and *global impact* (i.e., more than one interested country from different regions).

It should be noted that there can be different levels of "global impact", depending on how many different locations show interest in the event (e.g., a high-impact global world events will spark conversation in many countries). In addition, a location can be of any type of geopolitical division granularity, such as a city, a region, a country, a continent, etc. However, for our work we focus on locations at *country level*. Therefore, at times we use the concepts of "locations" and "countries" interchangeably. In particular, in the following section, we define a representation for relations among locations, which we exploit in Section 6 for extracting international relations.

3.2 Representing relations among locations

The spatio-temporal context-aware event representation allows us to extract different types of relationships among locations for a given event collection. In particular, we define a *protagonist-interest* vector **pi** for a location l_j, which represents the interest that other locations have in events that have l_j as a protagonist. We define **pi** for l_j as:

$$\mathbf{pi}(l_j) = \left[w(l_j, l_1), w(l_j, l_2), \ldots, w(l_j, l_{|L|})\right], \tag{2}$$

where,

$$w(l_j, l_k) = f(\#\text{ of events that have } l_j \text{ as protagonist in which } l_k \text{ shows interest}),$$
$$\forall l_j, l_k \in L. \tag{3}$$

Likewise, we also define the *co-protagonist* vector **cp** for the location l_j as follows:

$$\mathbf{cp}(l_j) = \left[w'(l_j, l_1), w'(l_j, l_2), \ldots, w'(l_j, l_{|L|})\right], \tag{4}$$

where

$$w'(l_j, l_k) = f(\#\text{ of events } l_j \text{ as protagonist in which } l_k \text{ is also a protagonist}),$$
$$\forall l_j, l_k \in L. \tag{5}$$

The relationships between locations, given by **pi** and **cp**, allows us to identify similarity relationships among locations, such as:

- **Locations that produce similar interest:** from **pi** we can extract sets of locations (countries) that are similar, based on the level of interest that they produce in other locations (countries). For example, they can be obtained using k nearest neighbors or by clustering locations' **pi** vectors.
- **Locations that are protagonists of the same events:** from **cp** we can identify which locations (countries) are similar, based on their interactions (i.e., they are protagonists of the same event) with other locations (countries). For example, they can be obtained using k nearest neighbors or by clustering locations' **cp** vectors.

The weights, $w(l_j, l_k)$ and $w'(l_j, l_k)$, are expressed as a function $f(x_{j,k})$, where $x_{j,k}$ corresponds to *# of events in which l_j and l_k interact*. In particular, for our visual tool described in Section 4, users have expressed the preference of visualizing the *absolute* number of events in which two countries interact (i.e., $f(x_{j,k}) = x_{j,k}$). Nevertheless, there are other cases in which the analyst could prefer the weights to reflect the fraction of events in which two countries interact in relation to the total of events for one of the two locations (e.g., $f(x_{j,k}) = x_{j,k} / \max(\#\text{ of events in which } l_j \text{ or } l_k \text{ participate})$). This can be useful in cases that the number of events in which different locations participate are very concentrated on specific locations. We explore cases such as these in Section 6, where we analyze the dataset described in Section 5.1 which is biased towards certain countries.

We note that weights can also be expressed as functions of the *# of tweets* or the *# of users*, and in addition, the proposed representation allows us to also specify *interest-interest* and *interest-protagonist* vectors, in a similar fashion to **pi** and **cp**. However, we do not focus on these variations at this moment.

Precision and recall of event locations. Empirically, we observed that the precision and recall of the locations considered to be protagonist of an event depends mostly on a ratio, which we call α. For an event E that contained more than one location, we defined α as the minimum percentage of tweets that must refer to a location l_i in relation to the most mentioned location $l_{\max}$, in order for l_i to be included in **P** or **I** vectors. Figure 1 shows an

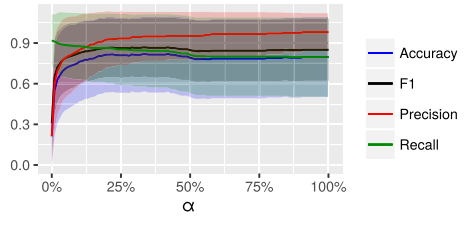


Figure 1 Mean and standard deviation of multi-label scores of accuracy, precision, recall and F1 measure by α ratio for 100 randomly selected events from our dataset.

empirical analysis of the effect of α on the precision, recall and F1 metrics of protagonist locations on a sample of 100 events. Precision and recall were estimated based on a manual assessment of the protagonist locations of those events. Based on this variation α can be set as the value that provides the best tradeoff between F1 and recall (α = 19% in our experiment).

Next, in Section 4 we introduce our visual tool for event exploration using the representation for events and international relations that we have just presented. In addition, in Section 6 we illustrate the use of our approach by presenting an exploratory data analysis that leverages the information contained by the aforementioned representations, based on the data collected by the visual tool.

4 Visualization tool

We present Galean, our prototype of a Visual Analytics tool to explore and retrieve news events based on our proposed geo-temporal context-aware event representation. We present our system's interface and high-level architecture. We show the usefulness of our tool by presenting two case studies, and by evaluating its effectiveness for new Information Retrieval tasks, such as: retrieving events that have particular countries as protagonists, and following international relations among countries over time.

4.1 Interface design

Galean's interface design is based on the Visual Information-Seeking Mantra: Overview first, zoom and filter, then details-on-demand [52]. Its interface (Figure 2) is composed of three main components: (i) filters and search (Figure 2(a), top section); (ii) a list of events and the main map (b and c in the middle section of Figure 2); and (iii) the timeline (Figure 2(d), at the bottom). A video demonstration of this tool is available at https://vimeo.com/150260355. In addition, a prototype of Galean focused only on Chilean news is available at http://galean.cl. In the future, the international version of Galean will be made available in the same location as the Chilean version. Next, we describe the interface and its components in detail.

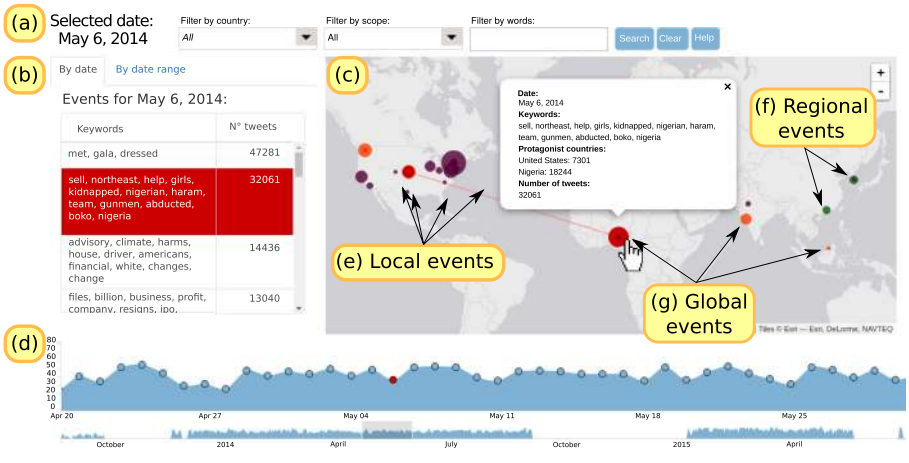


Figure 2 Galean overview. (a) Filters and keyword search options are in the top section. In the middle section, **(b)** a list of events by date and date range, and **(c)** the main map. **(d)** The timeline at the bottom shows the volume of news events over time.

Overview first: main map and timeline. The main map, table of news events and timeline provide a simple overview of thousands of tweets about news events. The main map shows events in their geopolitical context, represented as bubbles placed over the country or countries of their provenance. If the event is located in a particular city within the country, the bubble is placed in the city. On the other hand, if only country level information is available for the event, the bubble is placed on the country's capital. The size of each bubble represents the relevance of the event, measured by the volume of tweets associated with it. Purple bubbles (Figure 2(e)) represent events that are of local provenance (i.e., events in which only one country is the protagonist). Green bubbles (Figure 2(f)) represent regional provenance (i.e., more than one country is involved in the event, but all of them correspond to the same continent). Orange bubbles (Figure 2(g)) represent events which are of global provenance (i.e., more than one country is involved in the event and they belong to at least two different continents). If the cursor is placed over a bubble, a pop-up appears with information about the event. When the cursor is placed over green or orange bubbles - regional or global events - links appear to indicate the other countries that are relevant for that event. For example, in Figure 2 we observe several local events in the United States in May 6, 2014, indicated by purple bubbles located on this geographical area. In particular, the event with the highest impact is located in the West Coast. Some regional events (green bubbles) are located in South Korea and Brunei, and some global events (orange bubbles) are located in India, China and the United States. In addition, we highlight a global event that links the United States and Nigeria, which corresponds to the United States' intentions to send aid to Nigeria in response to the kidnaping of a large group of schoolgirls claimed by Boko Haram, in 2014.

To the left of the main map, the interface contains a list of events displayed by their most representative keywords and number of tweets. The timeline at the bottom shows the overall distribution of events over time, providing a historical overview of events per date. It is built as a focus-plus-context component of all the news events from the database. If a date is selected, the main map is updated showing only the events of that day. The map and timeline were implemented using Leaflet [53] and D3.js [54].

Zoom and filter. If the top filters of the interface are applied, the map, the list of events and the timeline are updated according to these filters. Events can be filtered by (i) whether they have one or more protagonist country, (ii) the scope of their provenance (local, regional or global, defined in Section 3), and/or (iii) by keywords. In particular, if more than one protagonist country is selected then the system retrieves only events in which those countries interact. For example, we can explore how the relationship between the United States and Nigeria evolved over time, based on the schoolgirls kidnapping by selecting both countries in the country filter and the word "`boko`" in the search box. By manually inspecting some dates in the timeline, we can retrieve several events related to that topic.

Details on demand: selecting a news event. To inspect a particular news event in depth, the user can click on its corresponding bubble in the map or on the list of events that is displayed. When an event is selected, shown in Figure 3, the map is updated to show a choropleth of the geographical distribution of tweets according to the countries that display interest in the event (countries from which users post tweets about the event). The event's protagonist countries are highlighted with a darker outline. Additional information for the event can be found at the right-hand side of the map. This information consists of a general event summary and of event tweets, categorized by source (i.e., reg-

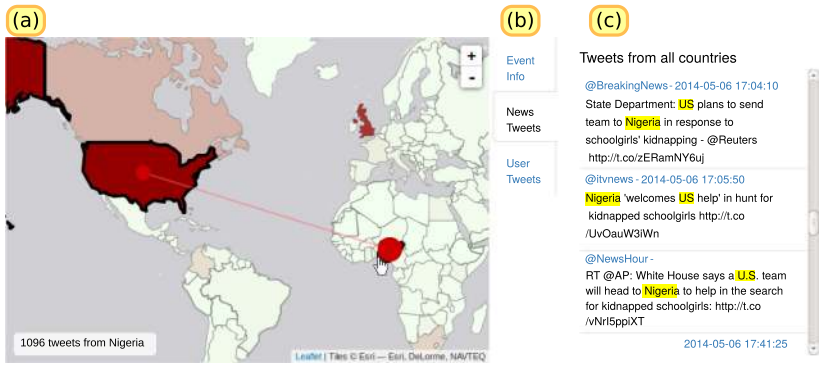


Figure 3 Details on demand for the news event about the intentions of the U.S. to send aid to Nigeria during the schoolgirls kidnapping (May 6, 2014). It shows the **(a)** geographical distribution of tweets, **(b)** additional information of the news topic divided into categories, and **(c)** tweets related to the event.

ular Twitter accounts or news outlet accounts), shown in Figure 3(b). By selecting these different sources, users can view a set of headlines for the event (i.e., when selecting news outlet tweets), or compare the people's perspective against that of the media. Finally, if a country is selected from the choropleth, tweets will be filtered to show only those coming from the selected country in chronological order. In future versions of the tool, we want to include an improved summarization technique to display tweets, such as organizing tweets by subtopics or a visual approach like ThemeRiver [55].

In particular, Figure 3 shows that most of the tweets related to the schoolgirls kidnaping come from the United States, Nigeria, Canada and Great Britain. In particular, the tweets shown in Figure 3 reflect the media's reaction to the event.

It is important to mention that our event exploration tool does not provide event ranking nor tweet ranking functionalities at the moment. The tool displays all of the events that match the user-defined spatio-temporal filters, and tweets are listed in chronological order. Ranking at the moment is not within the scope of our work, but it could be interesting to address in future versions.

4.2 System architecture

We present a general overview of the architecture for generating our event representations in order to use them in our application. The architecture, shown in Figure 4, consists of the following three parts: *"input"*, *"event representation generator"*, and *"visualization"*. The first component, (1) "input", is not part of our contribution, and is currently fulfilled by using existing methods, which can be replaced transparently as long as the requirements detailed next are met. On the other hand, the other two components, (2) "event representation generator" and (3) "visualization", are the core of our contribution and therefore essential to our system.

Given an input from the Twitter data stream we specify the following components of our framework (the particular setup for our proposed applications is detailed in Section 5.1):

1 **Input:** This module requires two subparts, the *"news event extractor"* and the *"geographical context extractor"*.
   (a) *News event extractor:* This submodule must output groups of tweets, where each group of tweets T should represent a cohesive news topic E. In particular, most of the tweets in the set T of an event E must be on the topic of a particular news

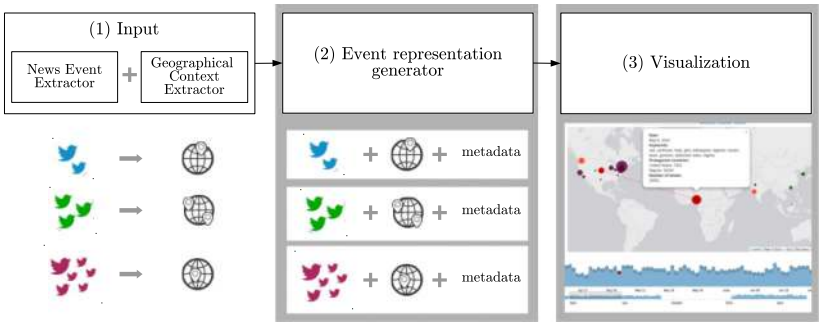


Figure 4 Framework consisting of three parts: (1) input, which collects data related to news event activity from social media and extracts its geographical information; (2) the event representation generator, which generates our representation of the input events and (3) the visualization, which consumes these events. Our contribution is related to the two latter modules, the first module can be replaced according to the task and/or state-of-the-art.

events. However, as we use a high-level representation of events, some noise is tolerated (i.e., tweets that do not correspond to the event).

(b) *Geographical context extractor:* This submodule associates spatial context to each tweet in T of each event E produced by the "news event extractor" module. Therefore, it must provide the geographical locations of the places mentioned in the text of the message and the geographical location of the author of the message (i.e., protagonist and interested locations, respectively). This module must locate the majority of the tweets in E correctly (i.e., with good precision) based on GPS coordinates and/or textual content, so that locations mentioned in tweets can be geotagged, and users can be geotagged as well (users can set their location using GPS coordinates or by using natural text).

2 **Event representation generator:** This component creates the event representations E for each of the groups of tweets provided by the "input" module. In particular this module must create a tuple E for each event, as specified by our definition in Section 3.1. This means that it has to produce the date D of the first tweet, a set of keywords K that describe the event, the set T of tweets and the $\mathbf{P}$ and $\mathbf{I}$ location vectors of the event.

3 **Visualization:** This module consumes the event representations produced by the "event representation generator" module and produces the event visualization interface.

5 Visual tool validation

We describe the validation of our visual tool, specifying our experimental setup, case studies, and user studies.

5.1 Empirical setup

We provide an overview of the data extraction methodology that we use for the *"input"* module in our architecture. The following modules are responsible for the creation of the input dataset from which the event representations are created in the following step. We emphasize, as mentioned in Section 4.2, that although the input data is important for the outcome of the final application, we consider the event detection and extraction to

be beyond the scope of our current work. In practice, this means that the way in which events are extracted can be replaced by another methodology in the future. However, at this moment we chose to rely on an existing approach that retrieves a set of events that are comprehensive and cohesive enough to test our system. Nevertheless, we acknowledge limitations in the type of events collected by our current setup, discussed in Section 7, but we believe that these limitations do not affect the generalization of the results of the proposed system.

News event extraction setup. The news event extraction module corresponds to that used by Kalyanam et al. [11], which consists of an ongoing process that periodically retrieves tweets about real-world news. We provide an overview of this process, which produces coherent sets of tweets about news topics, although with certain degree of noise that is well tolerated by our system. In particular, this is a two-stage iterative process that consists of (1) *news topic identification* (i.e., detection), and (2) *event tweet extraction*. We describe them briefly next (more details on this method, including the validation of the cohesiveness of the resulting events can be found in Kalyanam et al. [11]):

1 **Topic identification.** This approach does not detect events directly, but rather restricts itself to topics that have been posted on Twitter by mainstream news media accounts. The system periodically (each hour) retrieves headlines posted on Twitter by a set of *seed* news accounts, which must be provided. Using association rule analysis over the set of headlines collected in the cycle, the system outputs high-support sets of keywords ($\{K_1, K_2, \ldots, K_n\}$). These sets of keywords constitute terms that were posted together in a headline by more than one news outlet within an hour.

   In this particular setup, the seed set of news accounts correspond to 55 well-known international news media outlets (with verified accounts). These accounts are mostly from English-speaking sources based in the United States and Great Britain, such as @BreakingNews, @CNN, @NYTimes, @Jerusalem_Post, @AJEnglish, @NDTV, etc.[b]

2 **Data collection.** This stage iteratively takes the keyword-sets produced in (1), and uses each keyword-set $K \in \{K_1, K_2, \ldots, K_n\}$ to query the Twitter Search API in order to retrieve tweets T from *regular users* that also contain the keyword-set (i.e., that are commenting on the same news topic as the headlines). The search is done within the same hour in which the headlines were retrieved, removing tweets that where more than a few hours old, narrowing down the number of tweets that do not belong to the news topic due to the temporal relevance of the event. In principle, each keyword-set K is considered to be related to a unique news topic E. However, several keyword-sets could be referring to a same news topic (within one cycle or across several collection cycles), therefore an additional step is applied to merge one or more set of tweets into one within a one-day time window.

Geographical context extraction setup. We create a methodology for extracting the protagonist and interested locations, as well as their frequency for an event E with a set of tweets T. The toponym (i.e., location name) extraction and resolution phases are carried out using the off-the-shelf geoparser, CLAVIN [56]. However, since tweets are short and do not provide much context for toponym disambiguation, our methodology boosts the performance of the geoparser by adding context from other tweets in E.

As detailed in Section 3, this methodology relies in a ratio named α which we empirically set to 19% as that which provides the best tradeoff between F1 and recall, according to

Figure 1. This is, a location must have at least 19% of the mentions of the most frequent location of an event, to be considered as part of that event as well. Otherwise, we consider that this location was not actually involved in the event.

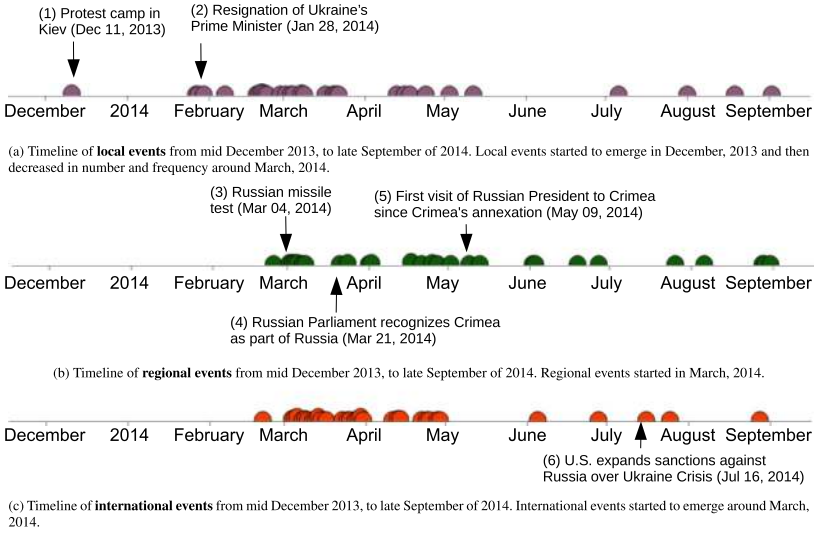
Dataset description. Using the previously described data extraction techniques, we collected a dataset of news events spanning from August 2013 to June 2015. This dataset consisted of 20,066 news events, which contained 193,445,734 tweets produced by 26,127,624 different users.[c]

We note that our event representation and applications are independent of the data extraction methodology. Therefore, in order to improve the representativeness of our event collection in the future, less biased methods of event extraction can be used, such as automatic event detection techniques [57, 58] and/or the integration of more comprehensive sets of seed news sources, as done for Chilean news analysis by Maldonado et al. [10].

5.2 Case studies

We use Galean to explore two selected news events: the *Ukrainian crisis*, dating approximately from November 2013 until today, and the *earthquake in Nepal* in April, 2015.

Ukrainian crisis. This event corresponds to the long-term conflict in Ukraine, which consensually started in November 2013 when the Ukrainian government decreed to suspend signing the "Association Agreement" [59] with the European Union. We used Galean to discover events related to the Ukrainian crisis, by selecting `Ukraine` in the country filter and the term `crisis` in the keyword filter. This retrieved only events that occurred in Ukraine and that contained social media messages with the term `crisis` between November, 2013 and March, 2015. To understand how local, regional and global events differ, we used Galean's filters to select the scope of each event. At the beginning (December, 2013), the majority of events were of local scope (Figure 5(a)), meaning that Ukraine was the only protagonist country at that time. Months later (March, 2014), regional and global events started to appear, indicating that other countries became involved in the cri-



(a) Timeline of **local events** from mid December 2013, to late September of 2014. Local events started to emerge in December, 2013 and then decreased in number and frequency around March, 2014.

(b) Timeline of **regional events** from mid December 2013, to late September of 2014. Regional events started in March, 2014.

(c) Timeline of **international events** from mid December 2013, to late September of 2014. International events started to emerge around March, 2014.

Figure 5 Timeline of local, regional and international events in the Ukrainian crisis between December, 2013 and September, 2014. Russia and the United States were the external countries that became the most involved in the Ukrainian crisis according to our analysis.

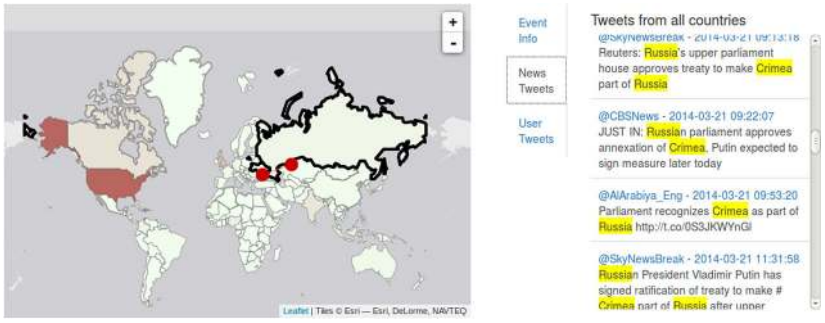


Figure 6 Russian Parliament recognizes Crimea as part of Russia (Point (4) in Figure 5(a)). Event detected on March 21, 2014. Total number of tweets: 7,660.

sis (Figure 5(b) and Figure 5(c)), tendency that started to decrease later in May, 2014. More precisely, Galean displayed 36 regional events about the Ukrainian crisis, 28 with Ukraine and Russia as protagonist countries. On the other hand, we found 48 global events, 12 of them involving only Ukraine and the United States, and 31 of them involving Ukraine, Russia and the United States.

In addition, we tracked some local events, in particular those related to the evolution of the protests in Kiev [60], and their consequences, such as the resignation of the Ukrainian Prime Minister at that time [61] (both events are marked in Figure 5(a) as `(1)` and `(2)`, respectively).

According to Galean, Russia and the United States were important actors in the Ukrainian crisis. Therefore, we explored more in detail some regional and global events. We found a series of events that related Russia to Ukraine, for example, when the Russian Parliament recognized Crimea as part of Russia in March 21, 2014 (Figure 6), and when, as consequence, the President of Russia, Vladimir Putin, celebrated Victory Day during his first visit to Crimea in May 9, 2014. Both events showed a strong impact on Twitter, with 7,660∼ and 11,647∼ related tweets.

Events that involved the United States included sanctions towards Russia [62], or accusations about Russia sending military help to separatists in Ukraine [63]. We used the filters provided by Galean to find relevant protagonist countries for certain events, and to track these events in time. For this case study, we observed an overall tendency of international, regional, and global scope events producing a greater impact, than local scope events.

2015 earthquake in Nepal. In this second case study we show how Galean can help users in crisis management, by looking at the causes of certain events. The starting point of this search was a news event about Japan signing an agreement to provide a loan for Nepal's earthquake recovery programs in December 2015 [64]. We retrieved events related to the earthquake by selecting `Nepal` as a protagonist country and `earthquake` as a keyword filter. In terms of scope, we obtained 24 local events, 7 regional events and 18 global events.

Regarding the earthquake's impact in social media, we observed that people's interest grew as the event evolved as evidenced by an increase in the number of related tweets and also of countries from which users displayed interest. In addition, we found emerging relationships between other countries, besides Nepal, such as the United States and India, as a consequence of having provided external aid for aftershocks.

Figure 7 Geographical distribution and sample tweets about the donor's conclave for the reconstruction in Nepal. Event detected on June 25, 2015. Total number of tweets: 2,565.

Given that our dataset extends only up to June, 2015, we were not able to follow the complete lifecycle of this event. The last global event related to the earthquake in Nepal stored in our database was from May 16, 2015, which accounted for the recovery of the bodies of the crew of the U.S. Marine chopper that went missing while helping victims [65]. However, after clearing the keyword filter in order to use only the filter by country, we found a regional event in June 25, 2015 about a donor's event among several countries to rebuild Nepal [66]. This event had Nepal and India as protagonists because the biggest donation came from India (Figure 7). Another agreement of this particular event was a loan from Japan to Nepal, which actually corresponds to the initial news that started this case study. Hence, by starting from that news, which is consequence of a past crisis situation, we were able to track its origin and subsequent events.

5.3 Expert feedback on the visual tool

We conducted a qualitative study of Galean with six domain experts using Pair Analytics [67]. Two specific aspects were investigated: (i) how intuitive and easy the tool was to use, and (ii) whether the tool could be used for the experts' day-to-day work in long-term news analysis. It is important to note that for this study our prototype only implemented two categories for provenance: local and international (regional was added afterwards). The international category included regional and global events.

Study design. Six users (two men and four women) were enrolled for the study, with ages ranging from 25 to 35 years. Four participants were journalists and the rest were people whose work relies heavily on news analysis. They were not economically compensated and participated voluntarily in the study.

After a detailed explanation of the goals of the study and a brief training session, the participants were asked to conduct three short tasks to test if they had understood the tool (e.g., *identify the date when most events happened*, *filter events by local and international impact and mention which of these scopes contains the most news events*). After those initial tasks were accomplished, participants were asked to carry out four more complex tasks, which aimed at more long-term news analysis. Two tasks focused on the exploratory capabilities of Galean and how it presents the evolution of news events over time. In those tasks, participants were asked questions such as *find news related to the Crimean crisis in 2014 and describe its evolution over time*. The last two tasks aimed at observing if users could discover patterns of news events and their propagation on Twitter. For those tasks,

users had to answer questions such as *how has the relationship between the United States and Iraq evolved over time in comparison to the relationship between the United States and Chile?* Finally, they were asked to discuss their experience using the tool.

Results. In terms of usability, all of the participants were able to complete all of the tasks without substantial problems and most agreed that with practice the tool was easy to use. Participants were able to track news over time, although it was not easy because the process was manual. Also, several participants said they enjoyed using the tool. In particular, participants were interested in exploring links between countries given international events and how the impact of news in social media changed over time. Two main usability issues were reported: to some participants it was not clear how the filter by date worked and some mentioned the clutter of events in the map when several news events were shown in the same location, even if strategies were applied to overcome the overlapping of bubbles.

Regarding the use of Galean for their daily tasks, all participants agreed that the tool was useful for analyzing news, however, most of the journalists indicated that this would depend of the type of analysis that they needed to perform. For most participants it was important to know the source that tweeted a particular piece of information, feature that Galean did not provide at the moment of the study. Indeed, a participant commented: "*without knowing the importance that a certain event had in news media, I can't say how much impact it had in Twitter.*" Given this feedback, in more recent versions of the tool we have included the name of the user who published the tweet and whether the user account corresponded to a known news outlet or not.

Two main patterns were mentioned by participants. The first, was that in general international events had more impact on Twitter than local events. The second was that some countries had more influence in social media than others. For example, a participant said: "*Well, the United States will always be under the magnifying glass [. . .] Remember that boy killed in Ferguson?* [68] *Everyone knew about that. But internationally, who remembers the boy killed by a policeman in Peñalolén (Santiago, Chile)?*"

5.4 User study

We conducted a more general user study to obtain evidence of users' perception of the visual tool, in relation to its efficiency and effectiveness for retrieving information about international relationships based on news reported on Twitter. As in the expert feedback evaluation section, we only divided events into local and international provenance scopes. We worked under two main hypotheses: **(H1)** Users will retrieve information about relationships between countries within the context of news events in a more efficient and effective way using Galean; and **(H2)** users will have a better subjective perception of Galean and a lower cognitive load when performing news event analysis.

Study design. The study had a within-subjects design, in which participants had to analyze news events using Galean, as well as using a competitive baseline interface based on SensePlace2 [29]. At the beginning of the session, the goals of the study were described to the participants and they were asked to fill a pre-study survey with demographic information. Next, the study was divided into two stages of news event analysis, each of them requiring participants to use one of the interfaces. At the start of each stage, participants followed a brief tutorial of the assigned interface and were given indications on how to complete the task. After they had finished, they were asked to fill the NASA Task Load Index [69] and a post-study survey of the assigned interface. Once they were ready, subjects

Figure 8 Baseline interface. The top section shows search by date and by text options. At the bottom left, it displays the tweets that matched the user search. Finally, at the bottom right it displays a map with the geographical distribution of tweets as a choropleth and the geographical entities that appear in the content of the tweets as bubbles over the resolved location. Both the choropleth and the bubbles representing a location can be used as filters for the tweets.

repeated the same procedure, with a different news event, with the second interface. We selected two news events for the users to analyze, and then asked questions about them such as "*when did the event happened?*" or "*which countries were involved in the event?*". To prevent a learning effect, we counterbalanced the order of presentation of each interface and of each event. In addition, the interface only gave access to tweets of one news event at a time.

All evaluations were conducted using the Chromium Web Browser in computers with an Intel Core i5 CPU, 8 GB of RAM and Ubuntu 14.04 installed. Participants spent close to one hour to complete the whole study.

Baseline. We built the baseline based on SensePlace2 [29], shown in Figure 8. We chose this tool as the most similar to ours in terms of the geographical information displayed. In the upper part, users were able to search by date and keywords. In the bottom-left, users could read tweets that matched the search. On the right side, the interface displayed geographical information in a similar fashion than SensePlace2, in which a map showed the number of tweets published by country and the geographical entities found in the content of the tweets. Since our focus is at country level, the geographical distribution of the tweets was not displayed as a grid, but only as a choropleth. The geographical entities found in tweets are represented as bubbles located in the geographical coordinates of the location. Both, the country area and the bubble on the map, could be used as filters.

Participants. Participants were recruited by e-mail and online forums in the Engineering School of the University of Chile. Given that our dataset was in English, we required them to have a good level of non-technical English. From the total of 30 participants recruited (3 of them were women), 5 of them were less than 20-years old, 20 were between 21 and 30-years, and 5 were between 31 and 40-years. In addition, 10 of them were undergraduate students, 8 were Masters students, and 12 were PhD students. Participants were not economically compensated, however refreshments were available during the study.

Results. Our study only partially supported hypothesis **H1**, evaluated by objective behavioral metrics of efficiency and effectiveness, but it completely supported **H2**, assessed by users' perception on the tasks performed during the study.

H1. *Objective measures of efficiency and effectiveness:* In terms of efficiency, users spent less time to complete the task using Galean (M = 895.58, SD = 317.9) than using the base-

line interface (M = 955.65, SD = 416.57), though this difference is not significant (p = 0.18). We argue that a reason for this difference being not significant is a learning effect, since some key components on the interfaces to complete the task were similar between conditions, such as the search box, the map, and the list of tweets. Therefore, we investigated this possible learning effect, and observed that users indeed spent less time using the second interface, but that this difference was more pronounced when Galean was second. By comparing the difference in time when Galean was the second interface ($p < 0.001$, Cohen's d = 0.74) versus when the baseline interface was used second (p = 0.013, Cohen's d = 0.6), we observed that the effect was larger when Galean was second. This result is interesting because Galean had additional components and interactions to learn from, which indicates that Galean was more efficient for this task than our baseline.

Regarding effectiveness, there was no clear difference in *recall* between Galean (M = 0.36, SD = 0.2) and the baseline (M = 0.35, SD = 0.2), p = 0.4, when used for retrieving countries. In terms of *precision*, Galean obtained a better performance (M = 0.952, SD = 0.11) than the baseline (M = 0.871, SD = 0.24), p = 0.062 when retrieving countries involved in a news event, although this difference was barely non-significant.

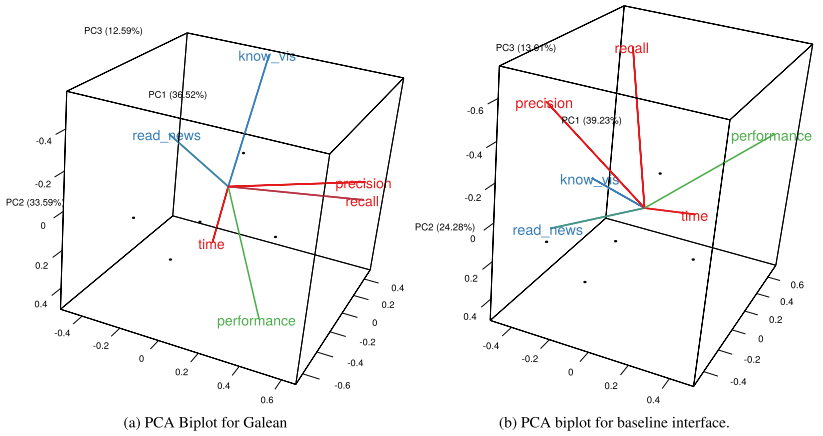
H2. *Subjects' perception on the interfaces.* Our study supported hypothesis H2, indicating that Galean was perceived in general as better than the baseline by users. We obtained subjective metrics by applying the NASA Task Load Index [69] and a post-study survey. Participants also showed the trend of requiring less effort to complete the task and less frustration ($p < 0.05$) when using Galean. With respect to the final post-study survey administered with a Likert 1-5 scale, people felt more confident about the information displayed in Galean than in the baseline ($p < 0.05$), they showed greater satisfaction ($p < 0.05$) and they were more likely to recommend it for eventual analysis of news events ($p < 0.05$).

User Agreement. In order to measure the level of agreement between users' perception over Galean versus the baseline interface, we used the Intraclass Correlation Coefficient (ICC) [70]. We calculated ICC between users (raters) over post-study survey questions (samples) and we report and interpret the values using the guidelines described by Koo et al. [71]. Values of ICC less than 0.5 are indicative of poor agreement, between 0.5 and 0.9 indicate moderate to good agreement, greater than 0.9 indicate excellent agreement.

The ICC results show a moderate to good level of agreement between users. For the case of Galean, the level of agreement was good (ICC = 0.887) with a 95% confidence interval from 0.722 to 0.977 ($F(6, 210) = 8.88$, $p < 0.001$). For the case of the baseline interface, the average measured ICC was moderate (ICC = 0.723) with a 95% confidence interval from 0.317 to 0.943 ($F(6, 210) = 3.61$, p = 0.002). ICC estimates and their 95% confident intervals were calculated using the *irr* package[d] version 0.84 within the R statistical package version 3.3.1 based on a mean-rating (k = 31), absolute-agreement and 2-way random-effects model.

Discussion. Our results show that in terms of user perception metrics, Galean clearly outperformed the baseline, but in terms of objective performance metrics, Galean shows only a tendency of better efficiency and effectiveness than the baseline.

To investigate these results further we conducted a principal components analysis (PCA) to integrate both the objective and subjective metrics (Figure 9) and we analyzed them by means of a biplot. A biplot is a projection-based graphical display which allows us to analyze multivariate data [72]. The word "bi" refers to the joint display of both rows and columns of an original data matrix, which has been projected into a lower rank approxi-



(a) PCA Biplot for Galean

(b) PCA biplot for baseline interface.

Figure 9 3D biplots of the principal component analysis for objective and subjective metrics for both interfaces. Some metrics were removed for clarity. Subjects knowledge such as how familiar were participants with visualizations (know_vis) and how frequently they read news (read_news) are in blue. Objective metrics of time, precision and recall are in red. Subjective metrics of user's perceived performance are in green. While precision and recall are related to previous user knowledge on the baseline, loading on the same direction of PC1, in Galean precision and recall are more related to user's perceived performance and barely related to previous user knowledge.

mation with rank $n = 2$ (2D biplot) or $n = 3$ (3D biplot). In our case, rows are user subjects and columns are variables, such as precision, recall, or time spent on an interface. We obtain the rank two and rank three approximations of our original matrix via PCA. Biplots are used for multivariate data analysis in areas such as sociology [73], genetics [74] and bibliometrics [75]. The interpretation of biplot displays is demonstrated by Gabriel [72] and more recently by Greenacre [76]. For instance, the closer the angle between vectors in the biplot, the larger the correlation between the variables represented by the vectors.

From this analysis we highlight two main results which support this discussion. The first is that for Galean, the subjective and objective metrics of performance were more consistent than for the baseline. Indeed, we observe in Figure 9 that precision and recall are closer to each other (in terms of angle between the vectors) and to the question about performance in TLX for Galean than for the baseline. Secondly, in the biplot for the baseline we observe that the variables 'familiarity with visualizations' (know_vis) and 'how frequently they read news' (read_news) are closer to the vectors of precision and recall and load in the same direction of the first principal component (the horizontal axis, which accounts for the larger variance in the data), which might indicate that previous knowledge of the users influenced their performance rather than the interface itself, though further analysis and a user study with a larger sample size are necessary to support this claim.

In summary, the additional evidence collected with both objective and subjective metrics indicates that Galean improves over a competitive baseline in several aspects.

6 Exploratory analysis

We present an exploratory data mining analysis that uses the information provided by our spatio-temporal context-aware event representation. We describe our empirical findings, which illustrate the usefulness of our proposed event representation. This analysis considers the location context of events at the country-level geopolitical division. This allows

us to explore the international interactions given by our current dataset. We note that the source code for this analysis along with additional information is available online [77].

The event extraction process, described in Section 5.1, was based on a seed set of internationally renowned news media accounts that publish information in English. This introduced a certain bias in our event collection towards events that took place in English speaking countries, and towards including more tweets in English than in other languages. For example, for the event *"correspondents dinner"* our current method will mostly retrieve tweets in English from users world-wide. On the other hand, an event described with a set of keywords which includes *"Barack Obama"* will retrieve tweets in several languages.

These biases must be taken into consideration because they can limit the representativeness of the findings yielded by our data mining analysis. Nevertheless, we believe that they do not invalidate our results, which show the perspective of a subset of the social network that is centered on news reported in the United States and Great Britain. Therefore, our results reflect the world-view of these two overly represented countries in particular, and of English-speaking users in general. Furthermore, other studies using the full Twitter stream, such as that of Poblete et al. [78], show a similar data distribution to ours, indicating that this type of bias could be inherent to Twitter itself.

Furthermore, an in-depth exploration of the bias in our dataset showed that the number of tweets produced during an event did not depend on the number of seed accounts that covered that event. Our analysis showed that only 13.5% of the users in the entire collection had actually reposted a tweet from the seed news media accounts, which gives the overall impression that these accounts did not influence much the amount of interest expressed by users. Also, we found no relation between the number seed accounts that shared an event and the number of countries that participated in the event in terms of provenance or of impact.

As mentioned in Section 3.2 we used a normalization for vectors **pi** and **cp**, defined in Equations 2, 3 and 4, 5, respectively. This normalization allows us to compare protagonist-interest and co-protagonist vectors in a way that mitigates the bias of overrepresented countries. In particular, for the **pi** vector we defined $w(l_j, l_k)$ as:

$$w(l_j, l_k) = f(x_{j,k}) = \frac{x_{j,k} - \mu(\mathbf{x}_{\cdot,k})}{\sigma(\mathbf{x}_{\cdot,k})}$$

and for **cp** we defined $w'(l_j, l_k)$ as:

$$w'(l_j, l_k) = f\left(x'_{j,k}\right) = \frac{x'_{j,k}}{x'_j},$$

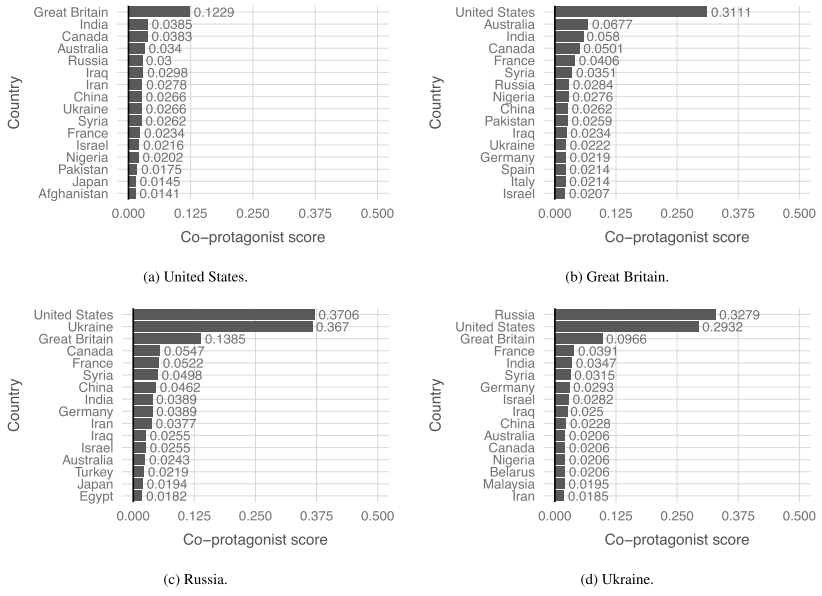
where $x_{j,k}$ was the number of events that have l_j as protagonist in which l_k is interested; $\mathbf{x}_{\cdot,k}$ is the vector containing the number of events in which location l_k is interested, $\forall l_j \in L$; μ and σ are the mean and standard deviation of the distribution of events, respectively; $x'_{j,k}$ is the number of events for which both l_j and l_k were protagonists, and x'_j is the number of events that had l_j as protagonist.

We started by characterizing the spatial distribution of our collection to describe its representativeness in terms of geographical coverage. In terms of protagonist locations, the United States and Great Britain were the protagonists of the majority of the events, followed by India, Australia, Ukraine and Russia (Figure 10(a)). The median number of

(a) The United States is a protagonist country of the majority of the events, followed by Great Britain, India, Ukraine, Russia, Australia, Syria and Iraq. A darker color corresponds to a higher level of protagonism.

(b) The United States is the country that showed the highest level of interest in other events, followed by Great Britain, India, Canada, Nigeria, Indonesia, Australia, and South Africa. A darker color corresponds to a higher level of interest.

Figure 10 Summary maps of interest and protagonists.



(a) United States.

(b) Great Britain.

(c) Russia.

(d) Ukraine.

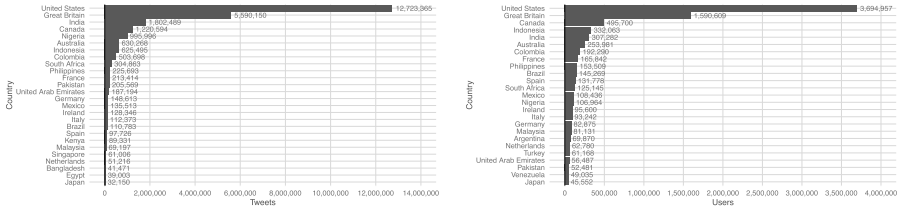
Figure 11 Relative co-protagonist measure of selected countries.

events in which countries were protagonist is 18.5, indicating that only a few countries were the protagonists of the majority of events. Figure 10(a) shows the distribution of the number of events in which countries were protagonists. When we computed the $\mathbf{cp}(c_i)$ vectors for selected c_i countries (Equation 4, normalized by the number of events in which a country c_i is protagonist), we observed that the United States and Great Britain were the protagonists of the majority of international events (Figure 11). There are some exceptions, such as Ukraine, which had only Russia as the co-protagonist many of its international events (Figure 11(d)).

In terms of worldwide interest, the countries that displayed interest in most events were the United States, Great Britain and India (Figure 10(b)). In addition, these countries also contributed the most tweets (Figure 12(a)).

We determined the location for 37.3% of the users (9,738,538 out of 26,127,625 users). These users were mostly distributed among the United States and Great Britain, followed by Canada, Indonesia and India (Figure 12(b)).

International relations exploration. We explored the dataset in order to identify similarity between countries according to the events in which they are co-protagonists and the interest shown towards these events by the rest of the countries in the world. We found that



(a) Tweets published per country.

(b) Unique users per country.

Figure 12 Description of the bias in the number of tweets and users, per country.

applying standard similarity metrics over the data in the event representations, yielded relationships between certain countries that resemble intense historical interactions and/or geographical proximity.

In terms of protagonist locations, we found countries that were *similar*, meaning that they were protagonists of the same events. In this case we used the Jaccard similarity between each pair of countries as our similarity measure, representing each country by the set of events in which it was a protagonist. The Jaccard similarity between two sets x and y is defined as $sim_{x,y} = \frac{|x \cap y|}{|x \cup y|}$. We filtered out the countries that were protagonists of less than 130 events (corresponding to the 80-th percentile of events for which countries were protagonist).

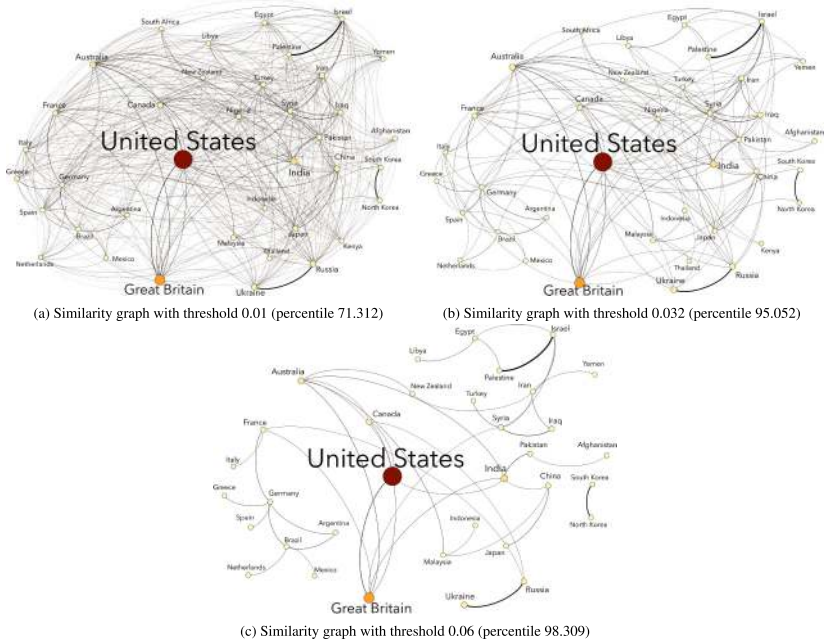
We studied the distribution of our similarity metric, in order to determine which relationships between countries were significant. We fitted the *similarity* to a theoretical probability distribution using the R package *fitdistrplus* [e] and we found that the best fit was a Gamma distribution with parameters *shape* = 0.8721 and *rate* = 85.7683. Based on this analysis, if S is a random variable with a Gamma distribution representing the similarities between countries, then we defined the similarity between two countries x and y as being *significant* if its value was in the 95-th percentile of the distribution, (i.e., if $P(S < sim_{x,y}) > 0.95$). Using this criteria, we determined a similarity threshold of $sim^* = 0.032$, above which we considered its value to be significant. This threshold can be parameterized at the 80-th, 90-th, or 99-th percentile, as the researcher finds appropriate. Table 1 shows the top-20 most similar countries based on this similarity, making it to the 97.181 percentile of our dataset.

We found that Israel and Palestine were the most similar countries, followed by Russia and Ukraine, North Korea and South Korea, Great Britain and the United States, and Iraq and Syria (Table 1). Their similarities are higher than the 99.25% of the pair-wise similarities in our dataset. There are real-world historical and geographical relations between those countries that can account for these similarities (for example, the Ukrainian crisis [79], or the Israeli-Palestinian conflict [80]). On the other hand, some of the similarities can be explained by the preponderance of certain events, such as the 2014 FIFA World Cup. These results indicate that there is information in Twitter data about real-world geo-political interactions which can be further studied using our event representation.

In Figure 13, we present three graphs where countries represent nodes and edges are weighted based on the Jaccard similarity. As we increase the threshold to connect two countries with an edge, communities of countries emerge. For example, in Figure 13(c), it is possible to identify a group consisting of Germany, Mexico, Brazil, Argentina, Nether-

Table 1 Most similar countries in terms of being protagonists of the same events (*co-protagonist* vector), using Jaccard Similarity. x'_i is the number of events in which country *i* was a protagonist

Country *i*	Country *j*	x'_i	x'_j	Similarity	Percentile
Israel	Palestine	561	360	0.2863	99.969
Russia	Ukraine	823	921	0.2094	99.906
North Korea	South Korea	158	179	0.1866	99.843
Great Britain	United States	4,015	10,162	0.0966	99.248
Iraq	Syria	654	647	0.0833	99.092
India	Pakistan	1,561	453	0.0753	98.998
Iran	Israel	496	561	0.0698	98.841
China	Japan	646	354	0.0605	98.340
France	Germany	627	371	0.0583	98.184
Argentina	Brazil	130	236	0.0578	98.152
Australia	Great Britain	974	4,015	0.0577	98.090
Brazil	Germany	236	371	0.0575	98.058
Syria	Turkey	647	198	0.0536	97.964
Iran	Iraq	496	654	0.0512	97.777
Australia	Malaysia	974	262	0.0492	97.682
Argentina	Germany	130	371	0.0481	97.620
Australia	India	974	1,561	0.0475	97.495
Germany	Greece	371	155	0.0457	97.401
Canada	Great Britain	715	4,015	0.0444	97.275
Egypt	Libya	316	253	0.0440	97.213
Great Britain	India	4,015	1,561	0.0436	97.181

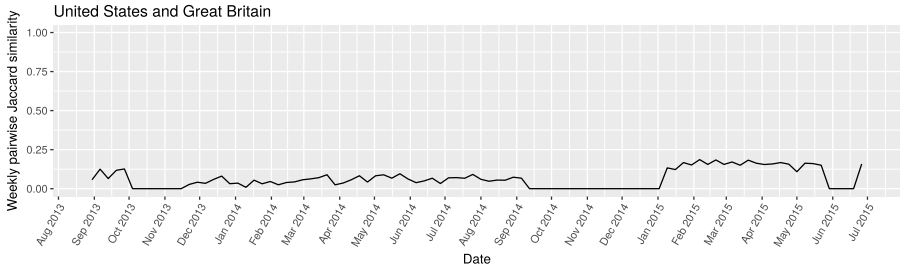


(a) Similarity graph with threshold 0.01 (percentile 71.312)

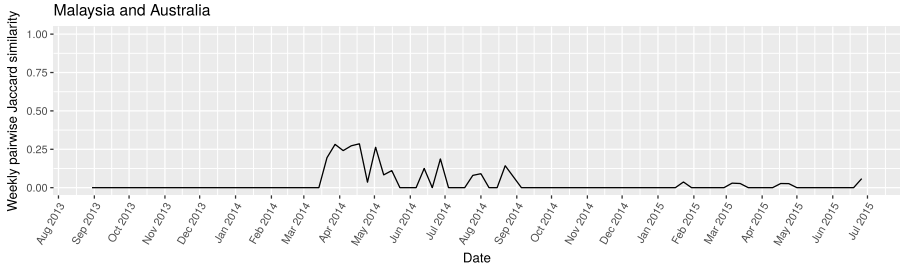
(b) Similarity graph with threshold 0.032 (percentile 95.052)

(c) Similarity graph with threshold 0.06 (percentile 98.309)

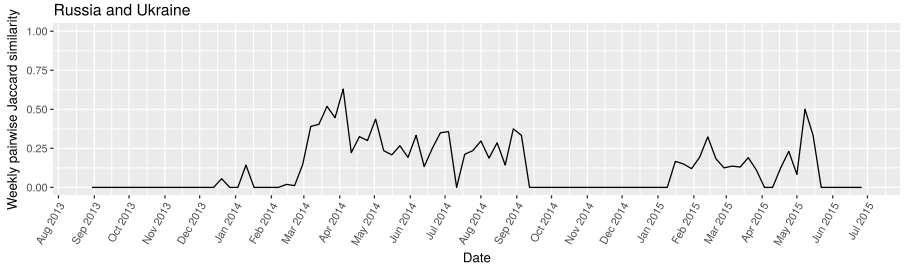
Figure 13 Similarity graphs of countries using the Jaccard similarity as the weight for the edges. Each node is a country and an edge between two nodes corresponds to the Jaccard similarity between those two countries. An edge is present if the similarity is higher than the given threshold. The node size and color represents the number of events in which each country was a protagonist, and the thickness of and edge represents the similarity.



(a) United States and Great Britain.



(b) Malaysia and Australia



(c) Russia and Ukraine.

Figure 14 Timeseries of the Jaccard similarity between co-protagonist vectors of selected pairs of countries over time. The value of similarity is computed for all the events in the given week. Data from October 2014 and December 2014 was not available.

lands, Spain and Italy: countries whose teams participated in the 2014 FIFA World Cup. Also, it is possible to observe edges among Malaysia, Indonesia, China and Australia, reflecting the disappearance of the Malaysia Airlines flight MH370 on 2014. Those two long-term events, for instance, sparked several events in our dataset, and the interactions between the protagonist countries are reflected in our analysis.

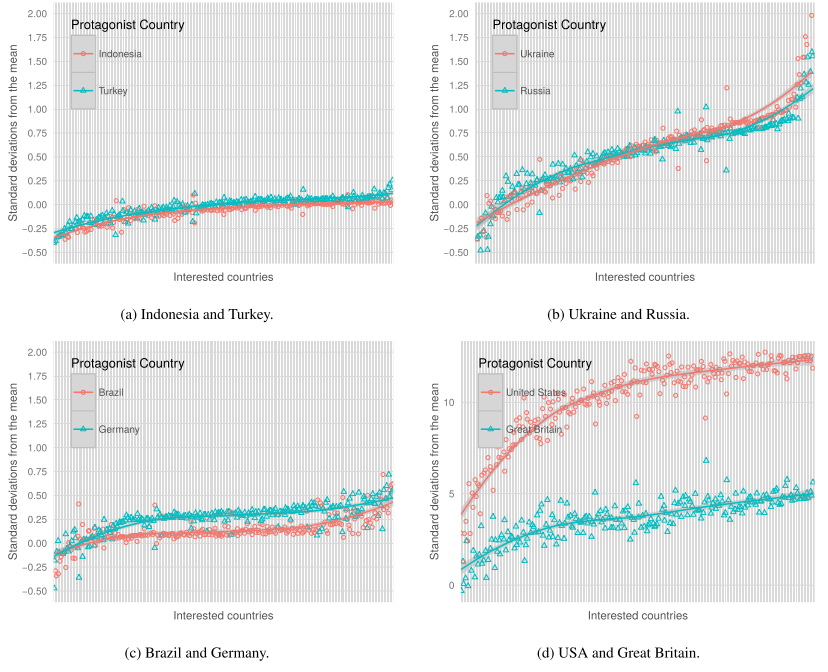
We further explored trends of co-protagonism by analyzing the similarity of countries over time. Given two countries, we computed their Jaccard similarity based on the events of a time window of one week. Figure 14 shows the time series between United States and Great Britain, Malaysia and Australia, and Russia and Ukraine. Each of those pairs of countries showed different characteristics in terms of how their similarity evolved over time. The US and Great Britain did not show notorious bursts of similarity over time, although they had high overall Jaccard similarity (Table 1), showing that although they were co-protagonists in several events, there was not a particular situation that suddenly increased their similarity in a narrow time span. On the other hand, Malaysia and Australia showed a burst starting in March 2014, shortly after the disappearance of the Malaysia Airlines flight MH370 (similar patterns arose when inspecting the relationship with Indonesia and China). Finally, Russia and Ukraine showed high values of similarity over time, start-

ing roughly in December 2013 and those patterns were maintained throughout 2014. This scenario correlates well with the case study reported in Section 5.2.

Another aspect that we explored was the interest that different countries had in events that occurred in different geographical regions. In other words, we explored the *protagonist-interest* relationship between countries. To do this, we represented each country c_i as its corresponding $\mathbf{pi}(c_i)$ vector (Equation 2).

We adjusted the original representation of the protagonist-interest vectors (Equation 2) in order to mitigate the data bias, which was reflected in that some countries were overly represented, because they produced much more tweets than others (Figure 12(a)). Hence, instead of counting the number of events with c_j as a protagonist, for which c_i expressed interest, we preferred to measure the interest of c_i in c_j as the difference between the average number of events of other countries in which c_i was interested, with respect to the number of events of c_j in which c_i was interested. In other words, our original interest measure was normalized by the average interest shown by c_i in other countries. Using this new interest measure we applied Euclidean distance to find the country c_2 with the closest **pi** vector to another country c_1 (Table 2). Given that there were countries that expressed interest in only a few events, or that were protagonists themselves of very few events, we only report the countries that were protagonists of at least 167 events (i.e., the average number protagonist events per country in our dataset). Figure 15 shows protagonist-interest plots for selected countries using the aforementioned measure.

We observed that Turkey had strong ties with other countries, being very close with several other countries according to protagonist-interest relations, such as Indonesia, Yemen, Afghanistan, Libya, and Malaysia. Furthermore, other similar countries were Italy and



(a) Indonesia and Turkey.

(b) Ukraine and Russia.

(c) Brazil and Germany.

(d) USA and Great Britain.

Figure 15 Protagonist-interest plots for selected countries. Each plot shows the level of interest (*y*-axis) displayed by the other counties of the world (listed along the *x*-axis) in the events of the featured pair of "protagonist countries". Country labels in the *x*-axis have been omitted for readability purposes.

Table 2 Pairs of countries that had the closest pi vectors according to the Euclidean Distance. x'_i is the number of events in which country *i* was a protagonist

Country *i*	Country *j*	x'_i	x'_j	Distance
Turkey	Indonesia	198	172	1.1442
Yemen	Turkey	202	198	1.3416
Afghanistan	Turkey	323	198	1.5304
Libya	Turkey	253	198	1.6050
Egypt	Palestine	316	360	1.6496
Malaysia	Turkey	262	198	1.8096
Japan	Spain	354	258	1.8327
Italy	Japan	315	354	1.9018
Brazil	Spain	236	258	1.9060
Germany	Pakistan	371	453	2.0674
Israel	Syria	561	647	2.4463
Russia	Ukraine	823	921	2.5557
Nigeria	Pakistan	412	453	2.5822
Canada	China	715	646	2.6025
Iran	Syria	496	647	2.6838
Iraq	Iran	654	496	2.9270
France	Canada	627	715	3.7859
Australia	France	974	627	4.1398
India	Australia	1,561	974	4.8339
Great Britain	India	4,015	1,561	41.7719

Japan, Brazil and Spain (and also Brazil and Germany); these similarities are be explained by the events triggered in the 2014 FIFA World Cup. Notably, Russia and Ukraine stand-out again, showing not only that they were protagonists of roughly the same events, but also that they were seen with similar interest by the rest of the world, making the impact that the Ukrainian crisis had on the news more evident. We also noted that most of these countries are close geographically, and as well as other countries, mostly from Asia. We argue that these results are another sign of the bias in our dataset: the perspective of international news as seen by English-speaking countries.

Finally, we explored events with the highest impact, considering international (Table 3) and local (Table 4) events. For this analysis, we considered all international events (regional and global). We counted the number of different interested locations for each event, however only considering interest measurements within the 99-th percentile of the dataset. From this analysis we were able to observe that the events with the highest overall impact covered several topics, and that the most recurrent events were sports and entertainment. Events like the death of the actor Robin Williams caused the most international impact, with a large number of tweets from 202 countries. This was followed by sports events, such as the 2014 FIFA World Cup, the 2013 Super Bowl and the boxing match between Floyd Mayweather and Manny Pacquiao. Other events with high impact included New Year's Eve for 2013, the *Charlie Hebdo* shooting in Paris, and the Grammy Awards in 2015. We also observed that the coverage of different news outlets was higher for these events. On the other hand, events with local impact consisted mostly of political events, such as political elections and debates, with the exception of a natural disaster and a sports event. We observed that in this case the coverage of different news sources was lower in relation to high impact international events, as well as the number of tweets involved.

Table 3 Events with most international impact, measured as the number of countries which showed interest higher than the 99-th percentile of overall interest

Event description	Tweets	Users	Outlets	Countries
Death of actor Robin Williams (2014)	1.8M	1.3M	48	202
FIFA World Cup final between Germany and Argentina (2014)	494K	385K	40	144
FIFA World Cup starts (2014)	476K	358K	45	143
Super Bowl starts (2015)	1.1M	849K	35	130
New Year's Eve (2013)	325K	279K	31	127
Soccer Player Luis Suarez is suspended from World Cup (2014)	213K	157K	38	106
Charlie Hebdo shooting in Paris (2015)	629K	328K	50	102
Grammy Awards (2015)	682K	432K	31	97
Boxing match between Mayweather and Pacquiao (2015)	779K	522K	37	97

Table 4 Events with most local impact, measured as the number of tweets coming from events with only one interested country, whose interest is higher than the 99-th percentile of overall interest. All events happened on 2015

Event description	Tweets	Distinct users	Outlets
US Supreme Court ruled in favor of same-sex marriage	51K	50K	7
Delhi Legislative Assembly election	35K	13K	3
Labour party said it will scrap the non-domiciled tax status	32K	15K	10
Tornado strikes Texas	31K	6K	4
TV appearance of Delhi chief minister candidate Arvind Kejriwal	30K	10K	1
Hillary Clinton announces presidential bid	30K	30K	3
Football player Cardale Jones announces he is returning to school	28K	22K	3

7 Known limitations

There are several limitations that we consider important to address. In particular, these limitations are not related to our proposed event representation, framework and implementation of our applications, but rather to the data extraction methodology which depends on external functionalities.

The news event extraction methodology relies on the headlines published by news media accounts. This technique provides good precision in terms of reporting events that did in fact exist in the real-world, but might omit informative events that did not receive media coverage. Therefore, the current data extraction approach can fail to retrieve events such as citizen movements and other important events that were informed only via social networks. In addition, as mentioned in Section 6, in the current data extraction setup the initial seeds for the event collection came from a reduced list of news media accounts, with limited country coverage and languages. Although the news event dataset likely represents a great majority of the news events and related tweets posted on Twitter, the collection will miss the long tail of events that had impact in other less represented countries worldwide. We note that there are several ways in which this bias can be mitigated in the future (see Section 6), all of them related to replacing external modules in the data input phase of the framework.

In addition, we note that although our proposed event representation can be considered generalizable to other social media platforms, we have not validated it on other sources of information besides Twitter. It is not certain that for other social media platforms we will have enough information, regarding user location and data availability, in order to produce accurate event representations.

Overall, basic future improvements of our work should consider:

- Implementing automatic event detection techniques for Twitter based on the data stream and network properties, as well as more comprehensive microblog event extraction approaches.
- Improving the geolocation tool accuracy. Despite CLAVIN's maturity as a geolocation tool, it does not recognize location names in languages other than English (even though the documentation of the tool indicates that it does recognize alternative location names [56]).
- Adding finer granularity to the geographical context extractor of our system, in order to include more precise administrative divisions such as cities and states.
- Merging events that discuss the same news topic in different languages. Recent approaches in cross-language microblogging retrieval [81] can be integrated for news event retrieval within our framework.

All of these improvements are however beyond the current scope of our work, which focuses on providing proof of the usefulness of the proposed event representation as well as the interactive user interface. Nevertheless, we are working on improving all of these features in future versions of our applications. For example, we have already started the task of providing more fine-grained locations for Chile and comprehensive sets of local news sources as in the work of Maldonado et al. [10].

Regarding our visualization tool, we note that even though it is an event retrieval tool, it does not focus on event ranking nor tweet ranking. At the moment the tool is centered on event exploration within spatio-temporal filters. In the future, event and tweet ranking functionalities could be added as optional features, incorporating state-of-the-art algorithms from these areas. So far, we have seen evidence that displaying the complete set of events, and tweets, by their chronological order, appears to be sufficient for event exploration.

8 Conclusions and future work

We have presented a spatio-temporal context-aware representation for news events in social media. Using this representation we have introduced a visual analytics tool named Galean that allows for retrospective analysis of real-world events through the aggregation of information posted by social media users. The main goal of our tool and our event representation is to allow exploration and quantitative analysis of events from a geographical and temporal perspective. In particular we introduce two types of geographical contexts for events: (1) protagonist locations, and (2) interested locations. The first corresponds to locations, in this case geopolitical divisions, that were involved in the event itself, and the second corresponds to locations where the event's information had the most impact.

Galean is designed to allow users to manually explore news events worldwide, as well as their impact and international relations implications. Using this tool, we show that the proposed event representation allows us to perform historical analysis of events and countries over time. Also, the visualization enables users to discover non-trivial information and patterns within events. To the best of our knowledge, this is the first tool that explicitly shows geopolitical links among locations given real-world events, allowing users to retrieve news by those relationships.

In addition, we introduce a quantitative data mining study over a 2-year Twitter dataset, in which we explore the properties of news events in social media and the international relations that are induced by those events. Our findings indicate that indeed there is new

information, which can be extracted at large scale, about how countries relate and how information is perceived in different places. Most interestingly, these relationships reflect historical relations that are found in the real world, indicating that there is value in social media data for historical research. Overall, our representation allows us to perform new IR tasks, related to exploring international relations and historical event retrieval.

In the future we will extend our representation to incorporate higher-level temporal properties of events, such as if the event is long-term, punctual or recurring, as defined in the work of Tan et al. [30]. Furthermore, we are interested in investigating how to automatically and visually support the discovery of cause/effect relationships over time for events. We are researching effective ways to display geographical and temporal evolution of variables (such as event impact) over time.

Another interesting line of research for future work is that of studying how pairwise similarities between countries evolve over time. This could be useful to automatically discover when new relationships arise and when deviations from normal patterns occur.

Finally, we are researching techniques to improve data extraction for news events, based on techniques such as those reported by Hasan et al. [82], to achieve a better representation of different locations, and to improve geolocation. Also, we are working towards creating automatic summaries of events in social media, on how to track and visualize event evolution over time, in particular the relations among protagonist countries that happen as consequence of these events, as well as incorporating approaches of cross-language information retrieval[83] for news event recommendations in microblogs.

Acknowledgements
Not applicable.

Funding
This work was partially founded by the Millennium Nucleus Center for Semantic Web Research under grant NC120004 and by CONICYT project FONDEF ID16I10222, grants PCHA/Doctorado Nacional 2013/21130470 (VPA), PCHA/Doctorado Nacional 2015/21151445 (MQ) and Fondecyt Iniciacion 2015/11150783 (DP).

Abbreviations
Not applicable.

Availability of data and materials
Data and its description are available at https://doi.org/10.6084/m9.figshare.5092678.v1.

Ethics approval and consent to participate
Written consent was provided by all users that participated in both our user studies, with full knowledge of how the data that they provided would be used. An ethics certificate was not required by the University of Chile since our studies did not involve sensitive user information.

Competing interests
No competing interests are declared.

Consent for publication
All authors have provided their consent for publication.

Authors' contributions
Conceptualization: VPA MQ BP. Data curation: MQ. Visualization Tool: VPA. Model Formalization: MQ BP. Investigation: VPA MQ. Methodology: BP DP. User studies: VPA DP. Software: VPA MQ. Data Analysis: MQ VPA. Project administration: BP. Funding acquisition: BP. Supervision: BP. Writing - original draft: VPA MQ BP DP. Writing - review & editing: BP DP. All authors read and approved the final manuscript.

Authors' information
Not applicable.

Author details
[1] Department of Computer Science, University of Chile, Avenida Beauchef 851, Santiago, Chile. [2] Department of Computer Science, Pontificia Universidad Católica de Chile, Av. Vicuña Mackenna 4860, Santiago, Chile.

Endnotes

[a] According to the division that considers 7 continents: Asia, Africa, Europe, North America, South America, Antarctica and Australia.

[b] Each Twitter account can be accessed in https://twitter.com/accountname, where `accountname` is the name of the account.

[c] This dataset will be available upon publication by contacting the authors, restricted by Twitter Terms of Services.

[d] https://cran.r-project.org/package=irr

[e] https://CRAN.R-project.org/package=fitdistrplus

Publisher's Note

Received: 30 November 2016 Accepted: 25 September 2017 Published online: 06 October 2017

References

1. Rogers R (2013) Debanalizing Twitter: the transformation of an object of study. In: Proceedings of the 5th annual ACM web science conference. WebSci '13. ACM, New York, pp 356-365
2. Inc., T. (2016). https://twitter.com. Accessed 22 November 2016
3. Wikipedia (2016) Comparative historical research. https://en.wikipedia.org/wiki/Comparative_historical_research. Accessed 22 November 2016
4. Castillo C, Mendoza M, Poblete B (2011) Information credibility on Twitter. In: Proceedings of the 20th international conference on world wide web. WWW '11. ACM, New York, pp 675-684
5. Sakaki T, Okazaki M, Matsuo Y (2013) Tweet analysis for real-time event detection and earthquake reporting system development. IEEE Trans Knowl Data Eng 25(4):919-931
6. Pak A, Paroubek P (2010) Twitter as a corpus for sentiment analysis and opinion mining. In: Proceedings of the seventh international conference on language resources and evaluation (LREC'10). European Language Resources Association (ELRA), Valletta
7. Saravanou A, Valkanas G, Gunopulos D, Andrienko G (2015) Twitter floods when it rains: a case study of the UK floods in early 2014. In: Proceedings of the 24th international conference on world wide web companion. WWW '15 companion. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, pp 1233-1238
8. Quezada M, Peña-Araya V, Poblete B (2015) Location-aware model for news events in social media. In: Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval. SIGIR '15. ACM, New York, pp 935-938
9. Peña-Araya V, Quezada M, Poblete B (2015) Galean: visualization of geolocated news events from social media. In: Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval. SIGIR '15. ACM, New York, pp 1041-1042
10. Maldonado J, Peña-Araya V, Poblete B (2015) Spatio and temporal characterization of Chilean news events in social media. In: TAIA '15
11. Kalyanam J, Quezada M, Poblete B, Lanckriet G (2016) Prediction and characterization of high-activity events in social media triggered by real-world news. PLoS ONE 11(12):1-13. doi:10.1371/journal.pone.0166694
12. Kamath KY, Caverlee J, Lee K, Cheng Z (2013) Spatio-temporal dynamics of online memes: a study of geo-tagged tweets. In: Proceedings of the 22Nd international conference on world wide web. WWW '13. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, pp 667-678
13. Leetaru K (2011) Culturomics 2.0: forecasting large-scale human behavior using global news media tone in time and space. First Monday 16(9)
14. Chakrabarti D, Punera K (2011) Event summarization using tweets. In: International AAAI conference on web and social media
15. Quezada M, Poblete B (2013) Understanding real-world events via multimedia summaries based on social indicators. In: Collaboration and technology. Springer, Berlin, pp 18-25
16. Alonso O, Bannur S, Khandelwal K, Kalyanaraman S (2015) The world conversation: web page metadata generation from social sources. In: Proceedings of the 24th international conference on world wide web. WWW '15 companion. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, pp 385-395
17. Wang X, Dou W, Ribarsky W, Skau D, Zhou MX (2012) Leadline: interactive visual analysis of text data through event identification and exploration. In: Proceedings of the 2012 IEEE conference on visual analytics science and technology (VAST). VAST '12. IEEE Comput. Soc., Washington, pp 93-102
18. Wang H, Can D, Kazemzadeh A, Bar F, Narayanan S (2012) A system for real-time Twitter sentiment analysis of 2012 U.S. presidential election cycle. In: Proceedings of the ACL 2012 system demonstrations. ACL '12. Association for Computational Linguistics, Stroudsburg, pp 115-120
19. Guille A, Favre C (2015) Event detection, tracking, and visualization in Twitter: a mention-anomaly-based approach. CoRR. 1505.05657
20. Ritter A, Mausam, Etzioni O, Clark S (2012) Open domain event extraction from Twitter. In: Proceedings of the 18th ACM SIGKDD international conference on knowledge discovery and data mining. KDD '12. ACM, New York, pp 1104-1112
21. Watanabe K, Ochi M, Okabe M, Onai RJ (2011) A real-time local-event detection system based on geolocation information propagated to microblogs. In: Proceedings of the 20th ACM international conference on information and knowledge management. CIKM '11. ACM, New York, pp 2541-2544
22. Abdelhaq H, Sengstock C, Gertz M (2013) EvenTweet: online localized event detection from Twitter. Proc VLDB Endow 6(12):1326-1329

23. Walther M, Kaisser M (2013) Geo-spatial event detection in the Twitter stream. In: Advances in information retrieval. Springer, Berlin, pp 356-367
24. Lee C-H, Yang H-C, Chien T-F, Wen W-S (2011) A novel approach for event detection by mining spatio-temporal information on microblogs. In: Advances in social networks analysis and mining (ASONAM), 2011 international conference on, pp 254-259
25. Krumm J, Horvitz E (2015) Eyewitness: identifying local events via space-time signals in Twitter feeds. In: Proceedings of the 23rd SIGSPATIAL international conference on advances in geographic information systems. GIS '15. ACM, New York, pp 20:1-20:10
26. Sankaranarayanan J, Samet H, Teitler BE, Lieberman MD, Sperling J (2009) TwitterStand: news in tweets. In: Proceedings of the 17th ACM SIGSPATIAL international conference on advances in geographic information systems. ACM, New York
27. De Longueville B, Smith RS, Luraschi G (2009) "OMG, from here, I can see the flames!": a use case of mining location based social networks to acquire spatio-temporal data on forest fires. In: Proceedings of the 2009 international workshop on location based social networks. LBSN '09. ACM, New York, pp 73-80
28. Dong X, Mavroeidis D, Calabrese F, Frossard P (2015) Multiscale event detection in social media. Data Min Knowl Discov 29(5):1374-1405
29. MacEachren AM, Jaiswal A, Robinson AC, Pezanowski S, Savelyev A, Mitra P, Zhang X, Blanford J (2011) Senseplace2: GeoTwitter analytics support for situational awareness. In: Visual analytics science and technology (VAST), 2011 IEEE conference on, pp 181-190
30. Tan Y, Vuran MC, Goddard S (2009) Spatio-temporal event model for cyber-physical systems. In: Distributed computing systems workshops, 2009. ICDCS workshops '09. 29th IEEE international conference on, pp 44-50
31. Lauw HW, Lim E-P, Pang H, Tan T-T (2010) Stevent: spatio-temporal event model for social network discovery. ACM Trans Inf Syst 28(3):15:1-15:32
32. Wikipedia (2016) Quantitative history. https://en.wikipedia.org/wiki/Quantitative_history. Accessed 22 November 2016
33. Michel J-B, Shen YK, Aiden AP, Veres A, Gray MK, Pickett JP, Hoiberg D, Clancy D, Norvig P, Orwant J, Pinker S, Nowak MA, Aiden EL (2011) Quantitative analysis of culture using millions of digitized books. Science 331(6014):176-182
34. Chadefaux T (2014) Early warning signals for war in the news. J Peace Res 51(1):5-18
35. Suchanek FM, Preda N (2014) Semantic culturomics. Proc VLDB Endow 7(12):1215-1218
36. DBpedia (2016) http://dbpedia.org. Accessed 22 November 2016
37. Huet T, Biega J, Suchanek FM (2013) Mining history with Le Monde. In: Proceedings of the 2013 workshop on automated knowledge base construction. AKBC '13. ACM, New York, pp 49-54
38. Robertson B (2009) "Fawcett": a toolkit to begin an historical semantic web. Digit Stud 1(2)
39. Meroño-Peñuela A, Ashkpour A, Van Erp M, Mandemakers K, Breure L, Scharnhorst A, Schlobach S, Van Harmelen F (2014) Semantic technologies for historical research: a survey. Semant Web 6(6):539-564
40. Marcus A, Bernstein MS, Badar O, Karger DR, Madden S, Miller RC (2011) Twitinfo: aggregating and visualizing microblogs for event exploration. In: Proceedings of the SIGCHI conference on human factors in computing systems. CHI '11. ACM, New York, pp 227-236
41. Jadhav A, Purohit H, Kapanipathi P, Anantharam P, Ranabahu AH, Nguyen V, Mendes PN, Smith AG, Cooney M, Sheth A (2010) Twitris 2.0: semantically empowered system for understanding perceptions from social data. In: Semantic web challenge, international semantic web conference (ISWC)
42. Purohit H, Sheth AP (2013) Twitris v3: from citizen sensing to analysis, coordination and action. In: Kiciman E, Ellison NB, Hogan B, Resnick P, Soboroff I (eds) ICWSM. AAAI Press, Menlo Park
43. Hassan S, Sanger J, Pernul G (2014) SoDA: dynamic visual analytics of big social data. In: Big data and smart computing (BIGCOMP), 2014 international conference on, pp 183-188
44. Bosch H, Thom D, Heimerl F, Püttmann E, Koch S, Krüger R, Wörner M, Ertl T (2013) Scatterblogs2: real-time monitoring of microblog messages through user-guided filtering. IEEE Trans Vis Comput Graph 19(12):2022-2031
45. Ertl T, Chae J, Maciejewski R, Bosch H, Thom D, Jang Y, Ebert DS (2012) Spatiotemporal social media analytics for abnormal event detection and examination using seasonal-trend decomposition. In: Proceedings of the 2012 IEEE conference on visual analytics science and technology (VAST). VAST '12. IEEE Comput. Soc., Washington, pp 143-152
46. Cao N, Lin YR, Sun X, Lazer D, Liu S, Qu H (2012) Whisper: tracing the spatiotemporal process of information diffusion in real time. IEEE Trans Vis Comput Graph 18(12):2649-2658. doi:10.1109/TVCG.2012.291
47. Dörk M, Carpendale S, Collins C, Williamson C (2008) Visgets: coordinated visualizations for web-based information exploration and discovery. IEEE Trans Vis Comput Graph 14(6):1205-1212
48. Global Voices (2016) http://globalvoicesonline.org/. Accessed 22 November 2016
49. Event Registry (2015) Event Registry system. http://eventregistry.org/. Accessed 22 August 2017
50. GDELT (2013-2014) The news co-ocurrence globe. Global Database of Events, Language. and Tone (GDELT) Project. Accessed 23 August 2017
51. BBC News (2014) Peru-Chile border defined by UN court at The Hague. http://www.bbc.co.uk/news/world-europe-25911867. Accessed 22 November 2016
52. Shneiderman B (1996) The eyes have it: a task by data type taxonomy for information visualizations. In: Proceedings 1996 IEEE symposium on visual languages, pp 336-343. doi:10.1109/VL.1996.545307
53. Leaflet (2015) http://leafletjs.com/. Accessed 22 November 2016
54. D3.js (2015) https://d3js.org/. Accessed 22 November 2016
55. Havre S, Hetzler B, Nowell L (2000) Themeriver: visualizing theme changes over time. In: Information visualization, 2000. InfoVis 2000. IEEE symposium on, pp 115-123
56. Berico Technologies (2012–2016) CLAVIN: Cartographic Location And Vicinity INdexer. http://clavin.bericotechnologies.com/. Accessed 22 November 2016
57. Metzler D, Cai C, Hovy E (2012) Structured event retrieval over microblog archives. In: Proceedings of the 2012 conference of the North American chapter of the Association for Computational Linguistics: human language technologies. NAACL HLT '12. Association for Computational Linguistics, Stroudsburg, pp 646-655
58. Choi J, Croft WB (2012) Temporal models for microblogs. In: Proceedings of the 21st. ACM international conference on information and knowledge management. CIKM '12. ACM, New York, pp 2491-2494

59. Aljazeera (2013) Ukraine drops EU plans and looks to Russia. http://www.aljazeera.com/news/europe/2013/11/ukraine-drops-eu-plans-looks-russia-20131121145417227621.html. Accessed 22 November 2016
60. Sky News (2013) Ukraine protesters now want leader's head. http://www.aljazeera.com/news/europe/2013/11/ukraine-drops-eu-plans-looks-russia-20131121145417227621.html. Accessed 22 November 2016
61. Yahoo News (2014) Ukraine PM resigns amid unrest, parliament revokes anti-protest laws. https://www.yahoo.com/news/ukraine-39-azarov-offers-resignation-government-press-083057414--sector.html. Accessed 22 November 2016
62. Fox News (2016) President Obama removing trade benefits for Russia over Ukraine. http://www.foxnews.com/politics/2014/05/07/president-obama-removing-trade-benefits-for-russia-over-ukraine.html. Accessed 22 November 2014
63. The New York Times (2014) Russia sent tanks to separatists in Ukraine. U.S. Says. http://www.nytimes.com/2014/06/14/world/europe/ukraine-claims-full-control-of-port-city-of-mariupol.html?_r=0. Accessed 22 November 2016
64. The Himalayan Times (2015) Japan assistance for Nepal earthquake recovery. http://thehimalayantimes.com/business/japan-assistance-for-nepal-quake-recovery/. Accessed 22 November 2016
65. The Himalayan Times (2015) All 8 bodies found at crashed US Marine chopper, Nepal army says. http://www.foxnews.com/world/2015/05/16/all-8-bodies-found-at-crashed-us-marine-chopper-nepal-army-says.html. Accessed 22 November 2016
66. Vatican Radio (2015) Donors pledge billions of dollars to rebuild Nepal. http://en.radiovaticana.va/news/2015/06/25/donors_pledge_billions_of_dollars_to_rebuild_nepal/1153906. Accessed 22 November 2016
67. Arias-Hernandez R, Kaastra LT, Green TM, Fisher B (2011) Pair analytics: capturing reasoning processes in collaborative visual analytics. In: 2011 44th Hawaii international conference on system sciences, pp 1-10. doi:10.1109/HICSS.2011.339
68. Fox 2 News (2014) Teenager shot, killed in Ferguson apartment complex. http://fox2now.com/2014/08/09/man-shot-killed-in-ferguson-apartment-complex/. Accessed 22 November 2016
69. Hart SG, Staveland LE (1988) Development of NASA-TLX (task load index): results of empirical and theoretical research. In: Hancock PA, Meshkati N (eds) Human mental workload. Advances in psychology, vol 52. North-Holland, Amsterdam, pp 139-183
70. Shrout PE, Fleiss JL (1979) Intraclass correlations: uses in assessing rater reliability. Psychol Bull 86(2):420
71. Koo TK, Li MY (2016) A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 15(2):155-163
72. Gabriel KR (1971) The biplot graphic display of matrices with application to principal component analysis. Biometrika 58(3):453-467
73. Galbraith J, Moustaki I, Bartholomew DJ, Steele F (2002) The analysis and interpretation of multivariate data for social scientists. CRC Press, Boca Raton
74. Yan W (2001) Ggebiplot - a windows application for graphical analysis of multienvironment trial data and other types of two-way data. Agron J 93(5):1111-1118
75. Torres-Salinas D, Robinson-García N, Jiménez-Contreras E, Herrera F, López-Cózar ED (2013) On the use of biplot analysis for multivariate bibliometric and scientific indicators. J Am Soc Inf Sci Technol 64(7):1468-1479
76. Greenacre MJ (2010) Biplots in practice. Fundacion BBVA
77. Quezada M (2016) Exploratory analysis. http://dcc.uchile.cl/~mquezada/galean/analysis.html. Accessed 22 November 2016
78. Poblete B, Garcia R, Mendoza M, Jaimes A (2011) Do all birds tweet the same?: characterizing Twitter around the world. In: Proceedings of the 20th. ACM international conference on information and knowledge management. CIKM '11. ACM, New York, pp 1025-1030. doi:10.1145/2063576.2063724
79. Wikipedia (2016) Ukranian crisis. https://en.wikipedia.org/wiki/Ukrainian_crisis. Accessed 22 November 2016
80. Wikipedia (2016) Israeli-Palestinian conflict. https://en.wikipedia.org/wiki/Israeli-Palestinian_conflict. Accessed 22 November 2016
81. Godavarthy A, Fang Y (2016) Cross-language microblog retrieval using latent semantic modeling. In: Proceedings of the 2016 ACM international conference on the theory of information retrieval. ICTIR '16. ACM, New York, pp 303-306
82. Hasan M, Orgun MA, Schwitter R A survey on real-time event detection from the Twitter data stream. J Inf Sci. doi:10.1177/0165551517698564
83. Grefenstette G (2012) Cross-language information retrieval, vol 2. Springer, Berlin