Eurographics Conference on Visualization (EuroVis) 2017
B. Kozlíková, T. Schreck, and T. Wischgoll (Editors)

Time-Ray Maps: Visualization of spatial and temporal evolution of news stories

J. Sheidin[1], J. Lanir[1], P. Bak[2], and T. Kuflik[1]

[1]University of Haifa, Israel
[2]IBM Watson Health, Foundational Innovation

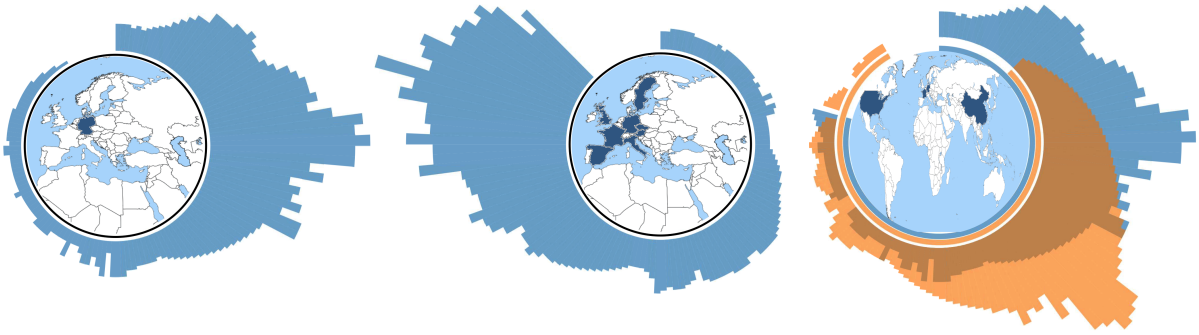

Figure 1: *Visualization highlights the temporal evolution and geographic origin of news stories*

Abstract

Time-Ray maps (see Figure 1) are designed to represent the temporal and spatial evolution of the reporting of a single news story. The visualization presents a compact representation that tells the story of the story in one glance. We show how our visualization can provide an imprint for different types of stories, enabling both comparison to other stories and means for analysis of the influence and evolution of a story. We demonstrate our technique on data coming from an online news aggregator platform. While specifically designed for news events, the method is also applicable for various types of time-series data that include intervening evolving events.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Picture/Image Generation—Line and curve generation

1. Introduction

News stories nowadays are presented by a multitude of news resources every day all over the world. These resources often cover the same topics and report redundant information. News aggregators and news portals such as Google News aim at gathering the important items from many sources and providing an interface to browse through the latest news items. However, they do not provide a way to analyze these items or to understand them in the context of time. For example, it is difficult to know for how long a single news item has been active, how important it is or what was its effect world-wide. The monitoring and analysis of news events is of interest to both professional users and to the general public. Thousands of news companies report huge amounts of news stories and events every day which is immediately available to users over the Web. Professional users, such as business or intelligence analysts are interested in the analysis of the evolution of these news stories to get valuable information and insights about trends in politics, business and public opinion around the world. In addition, analysis of the news reporting can be used in order to predict how a news item will continue to evolve. This can be of high interest to analysts, for example, in trying to predict whether a news story (e.g., a story about a protest) is fading or will continue to catch traction.

When analyzing a news story, one of the most interesting aspects is its temporal evolution - how does the story change and evolve over time. A news story can start unexpectedly such as in a report on an earthquake where the event starts with a peak without prior notice, or can gradually evolve, such as in a story about a political conflict. Similarly, a report on a news story can abruptly stop, for

DOI: 10.2312/eurovisshort.20171138

example in a report on a sports event that has just finished, or the story can gradually fade. Another aspect of a story that is of interest to analysts is the effect that a story may have. A news story can be "loud" and reverberate around the world creating multiple articles from multiple countries or can go relatively unnoticed and reported only locally, or worldwide but only on a small scale.

The current paper presents a novel visualization technique called Time-Ray map, that tells the story behind a single news story. Our objective was to create a technique that would enable analysis and understanding of both the temporal development as well as the geographical origins of the new reports conisiting of a story. Our visualization also supports the interleaving of several parts of the story we call *episodes*, showing the composition of the entire story, through the evolvement of each episode. While our visualization was specifically designed to support news stories, it can be applied to other data sources that may include multiple time series with spatial elements, such as in epidemic analysis, climate change and more.

2. Related Work

Much research has been done in the area of event and topic detection in text streams, with many works focusing on the detection and identification of news events [All12, SG10, APL98]. Probably the most common way to visualize a list of topics and their temporal change in their amount of influence is using the ThemeRiver (or stream graph) approach [HHWN02], which uses stacked graphs to show each stream's change over time. For example, in [DGWC10] and [MBB*11], topic streams were extracted from Twitter events and shown in a stream graph to show topical development. Similarly, in [GLYR07], top news stories were shown in a stream graph to show the thematic changes over time. The evolution of topical trends of news stories extracted from blogs has also been explored by using simple line graphs indicating term trends [FHRH08].

Other works take a different approach and visualize each story in a seperate line. In EventRiver [LYK*12], each stream of events is shown in a single, distinct line, with the changing vertical dimension representing the number of documents reporting or discussing it. Likewise, Leadline [DWS*12] automatically identifies meaningful events in news and social media data. Events, are visualized in a flow-like visual metaphor in which each stream (row) represents a topic and the height of each topic changes as the amount of textual information related to the topic varies. CloudLines [KBK11] also provide a row-based visualization, with each line representing a topic of interest (i.e., en entity). Events are plotted as points on a linear timeline. StoryTracker [KNAMK13] facilitates the analysis of similar topics that split and merge over time by showing the top stories in a column (rather than rows) for each time frame. The stories are then connected between each time frame to show how they relate over time. Similarly, [RBC*09] represents the changing nature of stories by clustering keywords into themes and then tracking their temporal evolution. Common to both the stream graph and the later approaches is that much of the focus is on the visualization of *multiple* streams or stories and the relations between the streams. On the contrary, our work focuses on the hour-to-hour development of a *single* story, providing a fingerprint of the story that could enable easy detection and understanding of the story development.

Other works, focused on other attributes of news events. For example, several works examined how the sentiment of news stories can be extracted and conveyed [WRM*09,DNKS10,LSB*16]. Others, focused on visualizing the video or image channels of news items [GLYR07]. Very few works looked at visualizing the spatial attributes of a news story. In [GHA*14], automaticly generated maps were created for a news story based on text mining and cartagrophy techniques. In [KBO*10], the source location of news stories was highlighted. However, they only looked at an entire news feed, and did not look at the change of reporting over time. Still, the challenge remains how to show both the spatial and the temporal aspects in a single view. In the current work, we examine the source locations of a single news story, allowing the user to investigate how the source locations, as well as how the magnitude of the story, change over different time periods of the story.

3. Method

We follow Munzner's nested model [Mun09] in our design. At first (*what?*) we describe and define data related to the news stories that are extracted from news feeds. Then, we describe typical user tasks (*why?*) that we aim at support. Third, we describe in detail *how* we encode the relevant information to visual attributes of the display.

3.1. News stories and episodes

A *news item* is a single news article reported in an online news source. Each news item includes a time-stamp that indicates when the news item has been published and a news source indicating who (and where from) reported the item. We define an *episode* as a series of related items similar to [KBK11]. In the news domain, several reports from different sources all reporting on the same news event comprise of a *news episode*. Finally, a *news story* can include one or more news episodes. For example, the Syrian civil war is a continuous news story in which specific conflicts or political events are considered as episodes. Our main focus has been on the temporal aspects of episodes, especially their development over a particular time period and the spatial location and extent originating the discussions. The development of an episode is assessed by the number of relating single news items that build up the story.

3.2. Exploring news stories

The user task we aim to support is the temporal and spatial exploration of news episodes and stories. The temporal evolution of a story can be of great interest to analysts as well as the general public. The number of reports corresponding to a story and its temporal evolution may indicate the story's popularity, duration and importance. Users want to distinguish between breaking news or slowly developing stories that remain in the news for long period of time. The spatial dimension can reveal to the user the intrest points of the story, weather a story is more local or has specific global political or economic aspects. Finally, the relation between the temporal and spatial evolution of a story may also reveal valuable insights.

3.3. Visual design

We have limited our design space to a radial temporal axis for several reasons: the field of view is better exploited in radial visu-

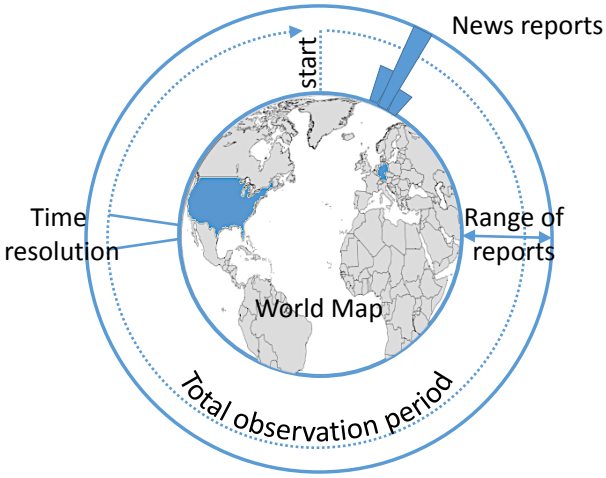

Figure 2: *Visual design showing the mapping of data attributes to visual properties of the display.*

alizations, and therefore the overall picture is easier to reach the viewer than in a sequential view, in which corners are harder to access [BBBD08]. From an aesthetic point of view, radial views are often associated with symmetry, wholeness and closure, which are important properties when comparing temporal developments [BBBD08]. Finally, a radial view allowed us to nicely add the spatial view and to easily link it with the sequential time information.

We were inspired by the Sunburst visualization in creating the radial axis. The observation period was mapped to the perimeter of the radial view. Since visual elements close to the center are given less display space than those close to the circle circumference which typically introduces a bias, we created an empty disk in the interior of the chart to lessen this bias. The empty interior of the chart was used to show the spatial information associated with the news reports. Thus, we positioned an interactive radial world-map into this space of the chart. In our examples, the map encodes the origins of news, however, one could show any spatial information using the map. Segments around the circle have the same shape but are differently oriented, which in our design was encoded by time units, similar to minutes and hours of a watch. In order to encode the number of news items, we placed bars with a computed height on the corresponding time segment of the radial charts. Figure 2 shows a summary of our visual design.

Prior to its creation, the time range of the entire visualization as well as each bar needs to be defined. For example, one could use the circle to depict 12 hours, 24 hours, or more than one day of a news story. Obviously, while comparing between two different stories, the same time range should be chosen. In addition, the start location of a story needs to be determined. The story can start corresponding to the time of day, with the hours corresponding to the circle times (i.e., 12:00 am at the top of the circle), or we can situate the beginning of the story to start at the top of the circle, regardless of the hour it happens. For example, we might wish to align stories on a timeline (taking the first approach) to see what happened during the day, or if we wish to compare between two or more stories regardless of when they happened during in the day, we can take the second approach and align them at the top. Time labels annotated on top of the visualization can be used to clear up ambiguities.

3.4. Interaction support

Interaction support is vital for users in order to explore the visualization and extract insights. In our design we considered the following interactive features:

Tooltip When hovering over a bar, a tooltip indicates the exact number of news reports reported in that bar and a representative headline is displayed.

Selection When clicking a bar of an episode, we select one of the news reports included in that bar, and open its web-page in a new browser tab.

Interactive world map Allows the user to zoom and pan to the regions of interest on the map. Since our design constrains us to use a circular map projection, omitting half of the world's surface and distorting regions on the perimeter of the visible area, interaction is indispensable for exploration.

Time-segment selection Users can select time segments of interest using mouse click (to select a single bar) and mouse drag (to select a segment) . Once a time segment is selected, linking and brushing is conducted to the map, highlighting the corresponding spatial information for that selection. Through this linking and brushing users are able to efficiently explore the news reports and gain deeper understanding of the evolvement of the spatial distribution.

3.5. Pattern exploration

Examples of possible patterns in artificial datasets are shown in Figure 1. Figure 1a shows a local news pattern (since only one country is marked on the map), in which news items increase quite quickly in a fairly short duration and then gradually fade. Such a pattern could be for example, with some local event, that does not reach beyond the borders of local news. The fast peak suggests an unexpected event. For example, this could represent a local celebrity scandal that just broke out. A comparable pattern, though with different orientation of the peak and shape, is shown in figure 1b., in which news items gradually increase, and peak towards the end of the time period, with a significantly faster decrease after reaching the peak. Such a pattern can be, for example, the primaries-election of a european country's leading party, since it is a known-in-advance event that peaks during the event itself. This example has Europe-wide consequences and therefore might raise interest in several of the EU member states. Finally, a story with two episodes is shown in the figure 1c., in which one discussion is replaced by another with similar extent.

4. Instantiation of the Method

We use data extracted from the European Media Monitor (EMM). EMM [SPVDG13] is a news aggregation system, that monitors over 3700 news sources collecting more than 100,000 news articles per day in 60 languages. Articles are clustered by EMM into stories that report about the same event. Each article is enriched with

various metadata, such as people, entities and organizations that are mentioned in the articles. We exatracted information on top news stories published at the EMM site for over six months using the website's RSS feed. For each story, we gathered all articles related to the story, including an exact publication time, so we could easily count the number of articles of each story in any time span. Each article also reports on its news source name and location.

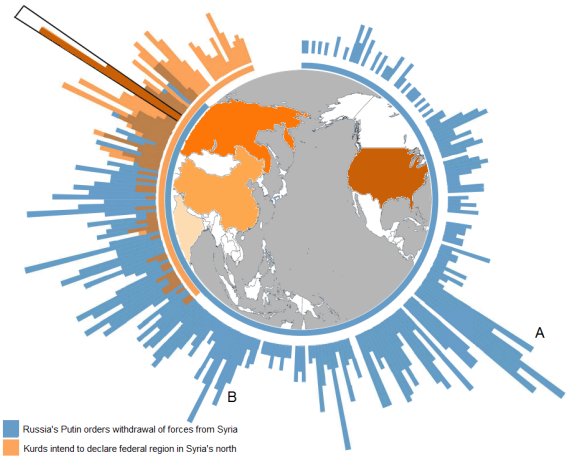

Figure 3: *Two episodes in the story of the Syrian war. Highlighted region show that the episode of "Kurds declaring federal region in the North" is increasingly discussed by Russian and US sources.*

Figure 3 describes the ongoing story of the war in Syria. We show two full episodes during a 48 hour span, with the beginning of the first story aligned at the top. The first episode, marked in blue, talks about the involvement of the Russian army in the Syrian war. Initial reports were scattered and talked mostly about the Russian support of president Asad, and the possible *intention* of Russia to pull out their troops. Since this was only speculations, we can see that the amount of reports was fairly low. After a while, we see a peak in the reporting (annotated as A). This peak corresponds with the Russian decleration of actually ordering the withdrawal of forces from Syria. This came at somewhat a surprise, and has many political implications and therefore received many reports during a relative long period (approximately 6 hours). We see a second peak starting later on (marked as B), which corresponds to reports on actual withdrawing of the planes. The second episode, marked in orange, reports on the Kurds intending to declare a federal region in north of Syria. While not too strongly related to the first episode, we can see the drift in world attention from one story to the other. Highlighting a small time area in this episode shows that this has been significantly discussed by Russian and US sources (and to a lesser extent by other countries). The visualization saliences the significant number of reports in that particular hour and the countries originating the sources in the interior map.

5. Reflections on design decisions

Our design space focused on a radial representation of time. While having the advantages of creating a perceptually concise and aesthetic representation, and enabling to easily combine the temporal and spatial features, this type of representation is not without limitations. The main disadvantage is with the ability of users to compare exact magnitudes. Users might have a hard time to judge proportions when bars are not aligned or have no leading baseline for comparison. Furthermore, the bars are wider as the distance from the center of the circle increases. The later bias can be overcome by scaling non-linearly the height of the units along the bar in distance from the center. We have applied tooltip support to provide accurate numeric information, However, we recognize that there is still much room for improving the readability of the Time-Ray chart when accurate information is to be retrieved.

Many of the related works have focused on visualizing various other attributes of news events such sentiment, social interaction, or topic related information. Further research can explore how the current technique can be altered to accommodate for such information. For example, in stories of a single episode or if episodes are aggregated, we can map sentiment to color, showing how sentiment has evolved over time in parallel to the magnitude and origin of the reports. In addition, our design does not consider the text itself. Future work would examine how to add automatic labels of important events tied to peaks or changes in the graph.

Overplotting occurs when events from different episodes of the same story are reported at the same time. We thought of two options to handle the overlapping of episodes. First, we can plot stacked bars, with each stack corresponding to one of the episodes. Alternatively, we can use slight opacity and indicate the overplotting with mixed colors, which has been our current choice. The choice between these two options might depend on the domain. In our case, rarely do more than two concurrent episodes occur. Swim-lanes and interaction (pressing a swim-lane can hide an episode) can provide further help. Nevertheless, this approach is not scalable in cases with increasing number of co-occurring episodes.

Finally, we used a choropleth map to indicate the amount of reporting per country. This choice might be biased since large countries, and countries with many news sources, such as the US, might have more sources reporting. One possible way to overcome this, is to normalize countries by population or the number of news agencies. For drawing the map inside the circle, we used an equal-area projection. Since the circle layout is used, it is hardly possible to show the whole world efficiently at once, which makes interaction (zoom and pan) indispensable.

6. Summary

The current work demonstrates a visualization technique to assess the geographic origin of news items simultaneously with their temporal evolution. Our novelty is in combining techniques into a single visual image that provides an informative and engaging view. We instantiated our method on a database of news sources, and reflected on our design decisions to show both advantages and shortcomings of our technique. Future work will compare Time-Ray maps to other linear time-based techniques, as well as expand the applicability of the technique to other domains.

References

[All12] ALLAN J.: *Topic detection and tracking: event-based information organization*, vol. 12. Springer Science & Business Media, 2012. 2

[APL98] ALLAN J., PAPKA R., LAVRENKO V.: On-line new event detection and tracking. In *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval* (1998), ACM, pp. 37–45. 2

[BBBD08] BURCH M., BOTT F., BECK F., DIEHL S.: Cartesian vs. radial–a comparative evaluation of two visualization tools. In *Advances in Visual Computing*. Springer, 2008, pp. 151–160. 3

[DGWC10] DÖRK M., GRUEN D., WILLIAMSON C., CARPENDALE S.: A visual backchannel for large-scale events. *Visualization and Computer Graphics, IEEE Transactions on 16*, 6 (2010), 1129–1138. 2

[DNKS10] DIAKOPOULOS N., NAAMAN M., KIVRAN-SWAINE F.: Diamonds in the rough: Social media visual analytics for journalistic inquiry. In *Visual Analytics Science and Technology (VAST), 2010 IEEE Symposium on* (2010), IEEE, pp. 115–122. 2

[DWS*12] DOU W., WANG X., SKAU D., RIBARSKY W., ZHOU M. X.: Leadline: Interactive visual analysis of text data through event identification and exploration. In *Visual Analytics Science and Technology (VAST), 2012 IEEE Conference on* (2012), IEEE, pp. 93–102. 2

[FHRH08] FISHER D., HOFF A., ROBERTSON G., HURST M.: Narratives: A visualization to track narrative events as they develop. In *Visual Analytics Science and Technology, 2008. VAST'08. IEEE Symposium on* (2008), IEEE, pp. 115–122. 2

[GHA*14] GAO T., HULLMAN J. R., ADAR E., HECHT B., DIAKOPOULOS N.: Newsviews: An automated pipeline for creating custom geovisualizations for news. In *Proceedings of the 32nd annual ACM conference on Human Factors in Computing Systems* (2014), ACM, pp. 3005–3014. 2

[GLYR07] GHONIEM M., LUO D., YANG J., RIBARSKY W.: Newslab: Exploratory broadcast news video analysis. In *Visual Analytics Science and Technology, 2007. VAST 2007. IEEE Symposium on* (2007), IEEE, pp. 123–130. 2

[HHWN02] HAVRE S., HETZLER E., WHITNEY P., NOWELL L.: Themeriver: Visualizing thematic changes in large document collections. *Visualization and Computer Graphics, IEEE Transactions on 8*, 1 (2002), 9–20. 2

[KBK11] KRSTAJIĆ M., BERTINI E., KEIM D. A.: Cloudlines: Compact display of event episodes in multiple time-series. *Visualization and Computer Graphics, IEEE Transactions on 17*, 12 (2011), 2432–2439. 2

[KBO*10] KRSTAJIC M., BAK P., OELKE D., KEIM D., ATKINSON M., RIBARSKY W.: Applied visual exploration on real-time news feeds using polarity and geo-spatial analysis. 2

[KNAMK13] KRSTAJIĆ M., NAJM-ARAGHI M., MANSMANN F., KEIM D. A.: Story tracker: Incremental visual text analytics of news story development. *Information Visualization 12*, 3-4 (2013), 308–323. 2

[LSB*16] LU Y., STEPTOE M., BURKE S., WANG H., TSAI J., DAVULCU H., MONTGOMERY D., CORMAN S. R., MACIEJEWSKI R.: Exploring evolving media discourse through event cueing. *IEEE Trans. Vis. Comput. Graph. 22*, 1 (2016), 220–229. URL: http://dx.doi.org/10.1109/TVCG.2015.2467991, doi:10.1109/TVCG.2015.2467991. 2

[LYK*12] LUO D., YANG J., KRSTAJIC M., RIBARSKY W., KEIM D.: Eventriver: Visually exploring text collections with temporal references. *Visualization and Computer Graphics, IEEE Transactions on 18*, 1 (2012), 93–105. 2

[MBB*11] MARCUS A., BERNSTEIN M. S., BADAR O., KARGER D. R., MADDEN S., MILLER R. C.: Twitinfo: aggregating and visualizing microblogs for event exploration. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (2011), ACM, pp. 227–236. 2

[Mun09] MUNZNER T.: A nested model for visualization design and validation. *Visualization and Computer Graphics, IEEE Transactions on 15*, 6 (2009), 921–928. 2

[RBC*09] ROSE S., BUTNER S., COWLEY W., GREGORY M., WALKER J.: Describing story evolution from dynamic information streams. In *Visual Analytics Science and Technology, 2009. VAST 2009. IEEE Symposium on* (2009), IEEE, pp. 99–106. 2

[SG10] SHAHAF D., GUESTRIN C.: Connecting the dots between news articles. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining* (2010), ACM, pp. 623–632. 2

[SPVDG13] STEINBERGER R., POULIQUEN B., VAN DER GOOT E.: An introduction to the europe media monitor family of applications. *arXiv preprint arXiv:1309.5290* (2013). 3

[WRM*09] WANNER F., ROHRDANTZ C., MANSMANN F., OELKE D., KEIM D. A.: Visual sentiment analysis of rss news feeds featuring the us presidential election in 2008. 2