Extracting and Visualizing Insights from Real-Time Conversations Around Public Presentations

Nicolas Garcia Belmonte

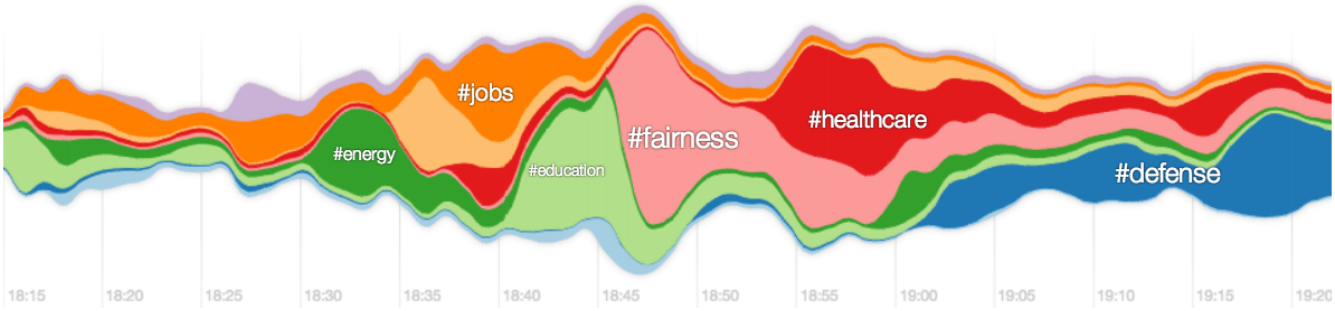

Figure 1: A streamgraph showing topic volume for Tweets sent during the State of the Union Address in 2014.

Abstract

During public addresses and debates (like the State of the Union Address) people engage in social networks generating thousands of messages. After the event happened, however, it is difficult to gather past conversations from those platforms and link them to the speech to relive the moment. To address this we developed a visualization technique that leverages the real-time conversational aspect of the Twitter platform to enrich the text of a public presentation with the discussion on Twitter during that speech; creating an annotated visual narrative of an event that can be replayed as the user scans the text of the speech.

Keywords: Text mining, real-time, streamgraph, Twitter, Tweet.

1 Introduction

During public addresses and debates (like the State of the Union Address) people engage in social networks generating thousands of messages. Whether it is politics, healthcare, jobs, entertainment, or education, people comment on these topics during public presentations. Twitter tracks these conversations on real-time, as the event unfolds.

After the event happened, it is especially hard to link comments made on social networks to the public presentation itself. Many news publications link some reactions on Twitter to special moments of an address, but this content is already digested and curated by someone. For these type of events curation is especially delicate since the main goal in general is to reflect as objectively as possible the public opinion.

To address this we developed an interactive visualization [1] that leverages the real-time conversational aspect of the Twitter platform to enrich the text of a public presentation with information discussed on Twitter during that speech. The visualization provides both an overview of the general topics discussed on Twitter during the event but also a detailed view of how the event was experienced paragraph by paragraph and minute by minute on Twitter. We will list some improvements that can be done to this approach and we will analyze how adaptive this technique is to other type of public presentations.

2 Data Collection

In order to create this visualization we collected two datasets: an annotated transcript of the speech with start and end timestamps for each paragraph, and the corpus of Tweets sent during that speech. We can filter Tweets by some hashtag (i.e keyword starting with "#") if there's a hashtag related to the event. In this example we worked with the 2014 State of the Union Address, and we filtered Tweets by the hashtag "#SOTU". We found thousands of Tweets linked to each paragraph. To provide a summary of all these Tweets we did a post-processing pass on the dataset to mine for topics discussed on Twitter during the address. With a set of keywords belonging to each topic we tagged each Tweet that contained one or more of those keywords with the right topic. Our final dataset consisted of Tweets tagged with topics, geo-location information and a link to the paragraph that corresponded (in time) with that Tweet.

3 Visualization

3.1 Visual Display

The visualization is an enriched view of the transcript for the state of the union address. The document is divided in three sections, top, right and center (figure 2). The center section contains the transcript for the public speech. When a paragraph in the transcript is set to "active", the paragraph is highlighted, and contextual information is shown for that paragraph in the other two sections. As the user scrolls down the page, the active paragraph changes to match the most visible paragraph in the screen.

The top visualization shows a streamgraph [3] that describes the volume of topics being discussed on Twitter during the address, minute by minute (figure 3). This view is a condensed summary of the overall discussion on Twitter during the address. This view serves another purpose as well: it is an index into the speech itself. By clicking on a specific part of the streamgraph the text scrolls to the paragraph that corresponds to that timestamp. This way, although we let the user explore linearly the address, we also provide her with a mechanism for finding fragments of the speech that would be more interesting to her. For example, if the user is interested in what was discussed around "jobs", she can click on the jobs section in the timeline and the text will scroll to that section of the speech.

The top right section is dynamic and provides information related to each "active" paragraph (figure 4). This section has two linked visualizations: a bar chart indicating the volume of Tweets

Figure 2: Page layout for the State of the Union address visualization. The visualization consists of two linked views that provide contextual information on the speech itself.

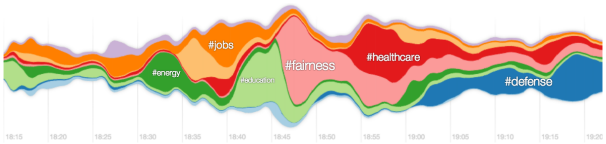

Figure 3: The Streamgraph visualization shows the evolution of the topics discussed on Twitter minute by minute during the SOTU.

around each topic for the active paragraph, and a choropleth map that shows for the selected topic in the bar chart its engagement distribution across the United States. The user can click on each bar of the chart to get engagement views for the selected topic in the choropleth map. The topic selected by default varies per paragraph and is the most discussed topic on Twitter for the current "active" paragraph.

3.2 Implementation

One of the main goals for this visualization is for the viewer to be able to share her findings with others easily through the internet. For this we used Web Standards to create the visualization. We use the JavaScript InfoVis Toolkit [5], a web standards data visualization framework to create the streamgraph visualization; we also use Data Driven Documents (D3) [2] for the choropleth and bar chart graphs. The former framework uses 2D Canvas to create the visualization; the latter uses SVG and HTML. We provided two mechanisms for sharing: when the user scrolls the page a permalink is generated for each active paragraph, so when the link to the

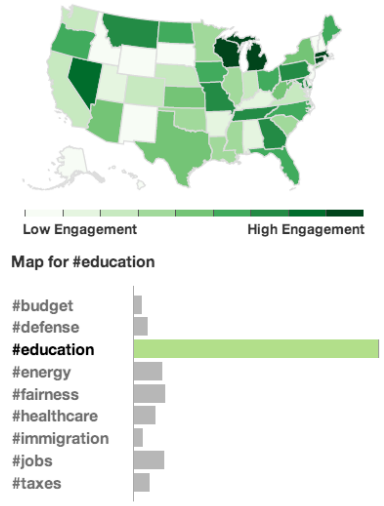

Figure 4: The bar chart and choropleth map show the real-time engagement distribution on Twitter for the highlighted paragraph.

visualization is shared it preserves the state of the page. We also made this visualization embeddable, and as a result news organizations like the Guardian [4], Mother Jones and others embedded the visualization on their site.

4 Conclusions and Future Work

We presented a technique that leverages the real-time conversational aspect of Twitter to provide context on what people discussed during a public discourse. Our specific example applies the visualization to the 2014 State of the Union Address; but by changing the map and topics this visualization can be extended to any form of speech and debate for the US and for any other country and region.

Some possible improvements on this visualization could be to add a view with actual quotes from Tweets sent for each paragraph. These Tweets could be chosen by defining a relevance metric based on number of retweets and favorites, whether the account is verified or not, location of the Tweet, etc.

Also it would be interesting to choose the topics in a way that not only show a correspondence between whats being presented in the speech and the real-time conversation, but also show discrepancies between whats happening on Twitter and the address. For example, if privacy was a topic which wasnt covered in the address, it would have been interesting to see if this lack of coverage initiated a conversation on Twitter about it.

Besides topics it would also be interesting to evaluate the general sentiment around each topic discussed on Twitter. For example, we could analyze if the conversation around defense during the SOTU was positive or negative. This can also be done by predefining a set of hashtags that can be used by Twitter users to express agreement or disagreement given a certain topic. This was used during various debates and interviews on TV around politics with high engagement results.

References

[1] http://twitter.github.io/interactive/sotu2014.
[2] M. Bostock. D3: Data driven documents. http://d3js.org.
[3] L. Byron and M. Wattenberg. Stacked graphs: Geometry & aesthetics. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1245–1252, Nov. 2008.
[4] M. Chalabi. http://goo.gl/jJ8xHQ.
[5] N. Garcia Belmonte. Javascript infovis toolkit. http://thejit.org.