ThemeStreams: Visualizing the Stream of Themes Discussed in Politics

Ork de Rooij
ISLA, University of Amsterdam
The Netherlands
orooij@uva.nl

Daan Odijk
ISLA, University of Amsterdam
The Netherlands
d.odijk@uva.nl

Maarten de Rijke
ISLA, University of Amsterdam
The Netherlands
derijke@uva.nl

ABSTRACT

The political landscape is fluid. Discussions are always ongoing and new "hot topics" continue to appear in the headlines. But what made people start talking about that topic? And who started it? Because of the speed at which discussions sometimes take place this can be difficult to track down. We describe ThemeStreams: a demonstrator that maps political discussions to themes and influencers and illustrate how this mapping is used in an interactive visualization that shows us which themes are being discussed, and that helps us answer the question "*Who put this issue on the map?*" in streams of political data.

Categories and Subject Descriptors

H.3.4 [**Information Storage and Retrieval**]: Systems and Software—*User profiles and alert services*

Keywords

Political tweets, visualization, real-time streams

1. INTRODUCTION

Over the past couple of years, politics and politicians have discovered social media as important means for communicating with voters and for influencing public opinion. Keeping track of the many discussion forums and other outlets is no trivial matter. What themes are being discussed? Who introduced a theme? Who "owns" it? Typical politically relevant themes include: the economy, healthcare, defense, foreign policy. According to a leading communication agency, during recent national elections in The Netherlands discussions revolved around approx. 500 issues, with differing levels and patterns of attention.

The participants of political discussions can often be mapped to a select number of so-called influencer groups. Specifically, one can identify the following four groups. First, there are those who currently actively have an (important) position within the governing body, the politicians. Second, there are those who lobby for (specific) important issues, the lobbyists. Third, there are journalists who specialize in politics as well as other high profile media influencers such as television stars or columnists. Fourth and fi-

nally, all other people taking part in political discussions we group together as the rest: the public.

In this demonstrator we describe ThemeStreams: an interactive visualization aimed at giving insight into the ownership and dynamics of themes being discussed, thereby enabling users to answer questions such as *Who put this issue on the map?* ThemeStreams allows users to explore streams of tweets, either from a fixed set of predefined themes or through a search box. It uses stream graphs [1] to indicated how influencer groups discuss a theme, thereby depicting the "aliveness" and ownership of a topic. Our visualization indicates when somebody said something, which influencer group this person belongs to, and it takes into account how many people react to a statement to estimate the "size" and "lifetime" of a statement. ThemeStreams can be accessed at `http://themestreams.xtas.net/`.

2. DATA AND BACKEND

For this demonstrator of ThemeStreams we focus on the Dutch political landscape. In the last couple of years Twitter has gained enormous popularity within Dutch politics, and is seen as an important outlet for talking about politics. Many politicians are actively using twitter and journalists regularly quote tweets in live television broadcasts. For this ThemeStreams demonstrator we therefore actively follow the following influencer groups:

- politicians (ministers, members of parliament, but also the local ranks of politicians in municipalities and provinces);
- journalists (news paper journalists as well as TV presenters);
- lobbyists (people pushing the people who are active in politics);
- other influencers (these include (satirical) columnists, politically engaged celebrities or stand-up comedians).

Tweets for the four influencer groups are being collected since late 2011. At the time of writing, the index driving ThemeStreams contains over 3.9M tweets. The four influencer groups together make up what we call the *inner circle*; it consists of about 1,400 individual users at the time of writing. In addition to the inner circle, we also consider and crawl a group of followers of members of the inner circle, approximately 245,000 users who together make up the *outer circle*. As we explain below, messages in the outer circle are used to weigh the impact of messages in the inner circle.

For content extraction and cleaning, we use the xTAS text analysis service.[1] For indexing, storage and retrieval we use Elastic-Search.[2] We use a hand-tuned stop word list that also filters most

---

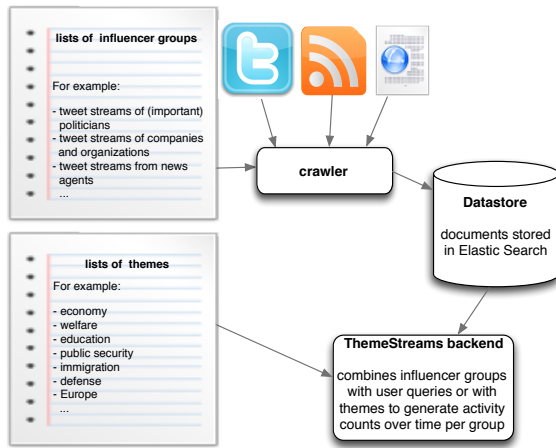[1] `http://xtas.net/`
[2] `http://www.elasticsearch.org/`

Figure 1: ThemeStreams backend.

common http link patterns so we only end up with words. Hashtags are stripped and indexed, and every mention is automatically selected as a candidate for inclusion in the outer circle. An overview of the ThemeStreams backend is given in Fig. 1.

3. INTERACTION AND VISUALIZATION

Users can gain insights in the development of messages around a topic in one of two ways. From the ThemeStreams home page they can access a fixed list of predefined themes and then explore streams of tweets around a theme they select. Alternatively, they can enter a topic in a search box (item A in Fig. 2).

In response to a topic submitted by a user (either predefined or ad-hoc), ThemeStreams displays a zoomable stream graph at the top of the page (item B in Fig. 2), depicting the number of tweets in the inner circle of four influencer groups retrieved for the topic. The thickness of the stream at each point in time is weighed by their "lifetime" (as determined by the number of retweets and mentions these tweets have received). In this way, we provide insight into how influential a group has been throughout the development of a theme, who finds a particular theme important and who were the first to talk about a particular theme.

Users can dive into more detail by zooming in using the focus + context principle [2]. In part C in Fig. 2, users can select a specific temporal interval, for instance because they know about important events related to their topic or because they observe interesting phenomena in the zoomable stream graph in part B of the interface. This allows users to not only see how important a theme was for an influencer group, but also what words one group used that other groups did not. To provide context, the stream graph for the entire period is also visualized (in part C of the interface); this enables rapid re-inspection of time periods close to the current focus.

The user's selection (indicated with a grey area, see part C), triggers the following events in the interface: (1) the zoomable stream graph in part B is restricted to the selected period and (2) in part D a term cloud is generated based on the tweets in the selected period. We offer two types of term cloud visualization selectable through the buttons in part E: one with a separate cloud for each of the influencer groups (see Fig. 3, right-hand side) and one with a combined representation with different colors indicated which influencer group was most influential for the term shown (see Fig. 3, left-hand side).[3]

---

[3]In order to comply with the Twitter ToS, the publicly accessible version of the demonstrator does not give access to tweets from which term clouds are generated.
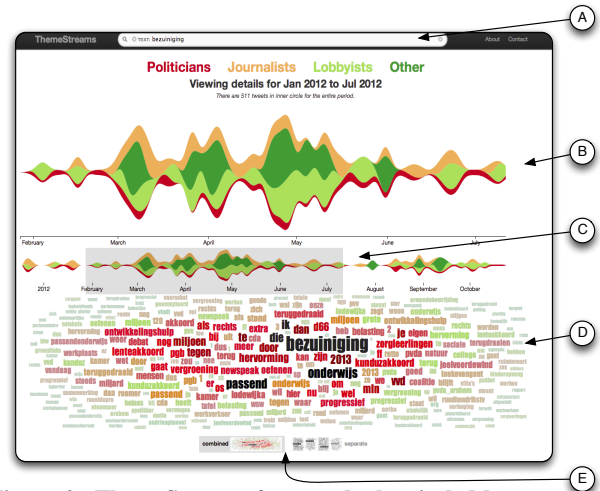

Figure 2: ThemeStreams front-end; the circled letters are explained in the text.

Initial usability studies have been carried out with colleagues and with professional end users, media analysts working for a communication agent. The main findings show that ThemeStreams was intuitive to understand, and inspection of parts of any query was easy to do. The combined cloud proved to be more insightful for fast overviews of the data. The individual clouds proved to be more useful for inspecting relative word usage between the groups. We also found a need for also depicting most represented speakers for any one group. A more detailed user study is currently in the works, and will be presented at a later time.

Figure 3: Users can switch between a combined cloud representation where color indicates originating influencer group, or separate representations per group, using option E in Fig. 2.

4. CONCLUSION

We have presented ThemeStreams, an interactive visualization for inspecting the political discourse over time using the Twitter posts of four political influencer groups. Other suitable application scenarios for ThemeStreams include discourse analysis over time in other domains, such as news paper archives.

5. REFERENCES

[1] L. Byron and M. Wattenberg. Stacked graphs–geometry & aesthetics. *Visual. and Comp. Graphics, IEEE Trans.*, 14(6):1245–1252, 2008.
[2] S. Card, J. Mackinlay, and B. Shneiderman. *Readings in information visualization: using vision to think*. Morgan Kaufmann, 1999.