Large Scale Human Sensing Over Time. Challenges and Lessons Learned

Omar Alonso
Microsoft
omalonso@microsoft.com

ABSTRACT

Twitter and Facebook continue to be top destinations for information consumption on the Internet. The ever-expanding social graph based enables the implementation of traditional features like item recommendation and selection of trending content that rely on human input and other behavioral data. However, given the enormous amount of human sensing in the world at any given moment in any platform, there is a lot of untapped potential that goes beyond simple applications on top of atomic level content like a post or tweet. In this talk we describe a social knowledge graph that discover relationships as they occur over time and how it can be used to capture the evolution of events or stories.

ACM Reference Format:
Omar Alonso. 2019. Large Scale Human Sensing Over Time. Challenges and Lessons Learned . In *Companion Proceedings of the 2019 World Wide Web Conference (WWW '19 Companion), May 13–17, 2019, San Francisco, CA, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3308560.3315001

1 INTRODUCTION

Social network posts reflect human sensing at scale. The magnitude of real-time information that is produced can be overwhelming for producers and consumers. If we just look at topics that have high activity volume, there is usually no context for the uninformed user and there is a quite a bit of redundant content for the informed user looking the latest update.

In general, trending topics and hashtags provide a very strong signal for important events that would later be covered in news articles. If the event is of interest to a wider audience, it is highly likely that it would eventually have its own Wikipedia entry. This information flow, from social posts to a final encyclopedia-like entry through a series of news articles is an emerging strategy for producing and consuming content. Current tools are not very well suited for discovering entities, summarizing events, or providing the evolution of a story and related topics. We are interested in mining social network data to construct new data that captures the dynamics and relationships of this large-scale human sensing as we have shown in our prior work [1, 2].

We describe infrastructure that supports querying and retrieving evolving stories about entities and events using a social knowledge graph, called SKG, derived from Twitter. The aim of the system is to fulfill information seeking scenarios by algorithmically generating the core of the story as it evolves over time by using selected relevant content derived from social data and news articles. This new data asset is a new type of document that is not as encyclopedic-centric like a Wikipedia entry but, instead, more dynamic to the many data components of the story, always up-to-date, and constructed in a fashion that allows different aggregations, applications, and, also very important, archiving.

Compared to previous work on knowledge base generation and information extraction from Twitter, we take a bottom-up approach with an emphasis on identifying good quality elements first. Selecting good content, users and links in an efficient manner from the Twitter firehose enables the creation of connections for high quality elements and populate the SKG schema (Figure 1) on a daily basis. The utility of SKG is to retrieve, extract, and present social information as a unit that can be beneficial to many applications.

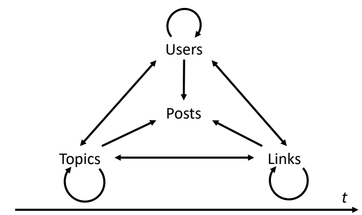

Figure 1: SKG schema: users, topics, links, posts as supporting evidence, and time.

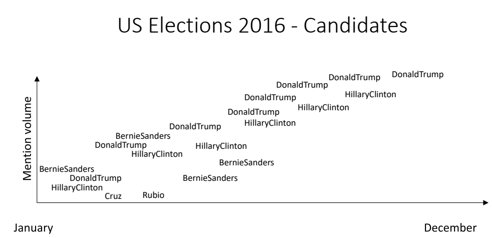

Figure 2: Visualization of the hashtags that represent candidates for the US Elections.

2 APPLICATIONS

We present a couple of applications. The first one, the evolution of an important event over time using different hashtags and the second, the construction of a Wiki-like document for a specific topic.

In the case of the 2016 US Elections, we can construct a scatter plot of hashtag frequency over time for the main candidates (both

WWW '19 Companion, May 13–17, 2019, San Francisco, CA, USA

ACM ISBN 978-1-4503-6675-5/19/05.
https://doi.org/10.1145/3308560.3315001

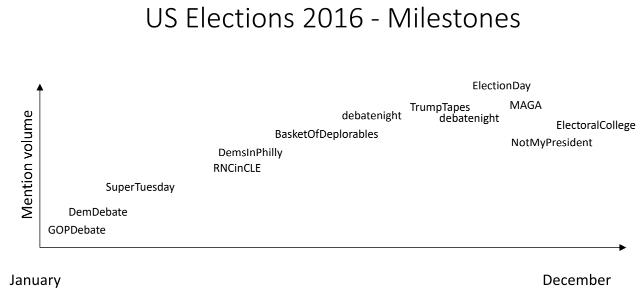

Figure 3: **Visualization of the milestones using hashtags that show main subevents (debates, famous phrase) that culminate in the election day in November.**

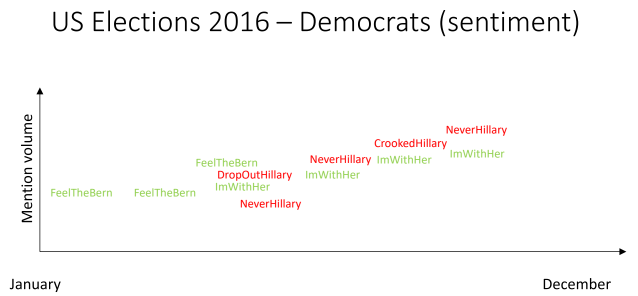

Figure 4: Visualization of the sentiment for Democrats over time. The colors green and red represent positive and negative sentiment respectively.

parties) as presented in Figure 2 for the whole year. We can observe that early in the year there was volume but not as much as we get closer to the elections, which is expected. Many candidates are mentioned in the first quarter of the year and then the two opponents dominate the rest of the chart.

Instead of using candidate names, we now produce a scatter plot of the most significant milestones (Figure 3). From the sequence of pre-established stages in the democratic process (e.g., primaries, debates, election day) to the final outcome (e.g., Electoral College), and the events that occurred within the year like Clinton's famous comment and Trump's tapes, the visualization clearly shows an overview of the elections.

We can also look at sentiment per party over time. For example, in the case of the Democratic party, a scatter plot of positive (`#ImWithHer`, `#FeelTheBern`) and negative (`#NeverHillary`, `#CrookedHillary`) hashtags is presented in Figure 4. Similarly, for the Republican party, a scatter plot of positive (`#MAGA`) and negative (`#NeverTrump`) hashtags is presented in Figure 5.

In Figure 6 we show the results of the Wiki-like generation for a topic as a query (`#ces2017`). Say that we are interested in getting an

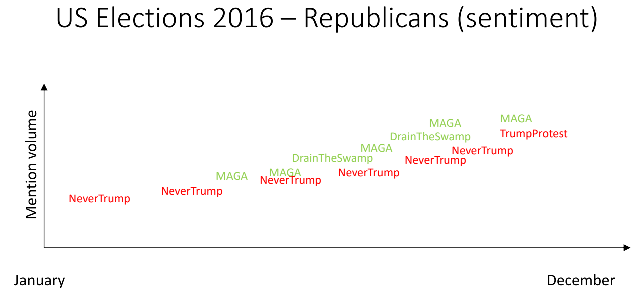

Figure 5: Visualization of the sentiment for Republicans over time. The colors green and red represent positive and negative sentiment respectively.

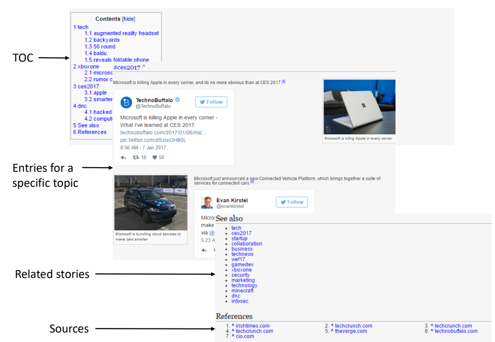

Figure 6: Wiki-like view for the topic `#ces2017`: table of contents, main story, related topics, and sources.

overview of what was discussed at the Consumer Electronics Show in 2017. The system queries SKG and returns a synthetic document that contains the typical elements of a Wikipedia page: table of content, story, related topics, and the sources that were used to create a story. We also include the original tweet as provenance.

3 CONCLUSION

There is little value at the atomic level when reading from a social networking firehose. Instead, we believe that new applications and aggregations can be implemented when looking at social network data as knowledge graphs. In contrast to knowledge-based solutions that retrieve facts, we are interested in stories and related subtopics. We described, at a very high level, a system that supports querying and retrieving evolving stories about events.

REFERENCES

[1] Omar Alonso, Vasileios Kandylas, and Serge-Eric Tremblay. 2018. How it Happened: Discovering and Archiving the Evolution of a Story Using Social Signals. In *Proceedings of JCDL*. 193–202.

[2] Omar Alonso, Vasileios Kandylas, Serge-Eric Tremblay, Jake M. Hofman, and Siddhartha Sen. 2017. What's Happening and What Happened: Searching the Social Web. In *Proceedings of WebSci*. 191–200.