

Does misery love company?

Change in sentiment over the Twitter social graph

COMP 4601 - Project

Student Name: Brian Ferch
Student Number: 100962115

Student Name: Jules Kuehn
Student Number: 100661464

Winter 2019

1 Abstract

(100-150 words) Describe main problem statement, goal, and scope of your project.

We investigate tweet sentiment over time for 400 socially connected Twitter users by dynamically visualizing changes in tweet sentiment for roughly 400,000 tweets from 2016 to present.

The project scope consists of crawling the social graph, gathering tweets for each user in our graph, translating and analyzing sentiment for each tweet, and visualizing these sentiments over time. Each user is represented as a node, and the color of each node indicates that user's sentiment at a given time. The graph is displayed using a force-directed layout. From an initial seed node, followers of a user are added along with directed edges for each (*user, follower*) directed relationship.

2 Introduction

Explain motivation for the project and briefly foreshadow your report, then introduce sections.

A 2015 paper by Tsugawa and Ohsaki [1] found that "negative messages are likely to be reposted more rapidly and frequently than positive and neutral messages". We did not investigate retweets but rather sought to answer whether sentiment spreads more generally

between users. This could confirm and expand on the results of that paper, or simply provide some useful insight into demonstrating the folk wisdom “misery loves company”: that negative sentiment is contagious.

In lieu of any substantial statistical analysis, we visualize the results on an animated graph, allowing for us to explore the data interactively and spot trends.

We will begin with a general overview of the data sources and technologies used and a review of related work. Then we will discuss the specifics of our implementation, and results obtained.

3 Background

Describe contextual knowledge a reader would require in order to comfortably understand your project report (ex. your data sources, libraries used, technologies discussed).

Twitter offers a free rate-limited API [2] which allows us to retrieve the required raw data from queries on a screen name or user id:

1. *userId* : a large integer, ex. 946341920036478976 for our seed user
2. *tweets* : most recent $\ell=3200$ tweets for this user consisting of $[(time_1, text_1), (time_2, text_2), \dots]$
3. *followers* : $[id_1, id_2, \dots]$

Many of the gathered users had predominantly non-English tweets. In order to run sentiment analysis with VADER (Valence Aware Dictionary and sEntiment Reasoner)[3] we first had to translate all tweets to English. This was accomplished using Google Cloud Translate [4], a commercial service.

Based on visual analysis of the data, cleaning was performed programatically. This yielded several datasets usable by the client:

1. *tweets* : all tweets for all users, sorted by time $[(time, userId, sentiment, text), \dots]$
2. *userInfo* : information for specific users such as average sentiment, number of tweets, and *followers*.

This data was then displayed as a graph in JavaScript using the d3.js library [5]. Sentiment is then animated on each node by the timestamps of tweets. Some smoothing (rolling window averages) is selectively applied to the data to aid in spotting trends at different time granularity.

4 Related work

Projects with similar scope, or technologies related to the implementation of your solution. Briefly mention the projects and their relation to your report, and annotate appropriately.

5 Methodology

Explain implementation decisions and details. Be specific but concise, referencing information about sources when required to explain functionality. What were your goals? Why did you implement the way you did - and how did you?

6 Discussion

This is the opportunity to showcase analysis or resulting benefits from your project implementation. Without getting into Future Work too much, briefly discuss what setbacks and desires your project implementation experience has left you with. Other than that not much advice on discussion; varies by project scope.

7 Future work

Discuss possibilities of future work, for example in terms of 1) what would you consider implementing given more time on this project and 2) how can this project be used by or furthered by the open source community to improve future works?

8 Conclusion

Summarize the content of the report with a focus on the goal, analysis, and resulting value. Recap relevant points from the discussion, but do not restate it. Finally, you can choose to touch on future works as you close out the paper.

A

Figures and tables

B

References