

# Final Project Proposal

## 1 Proposal

### (a) Social Media Analysis

Social networks contain an enormous amount of data. Facebook alone processes 2.5 billion pieces of content and more than 500 terabytes of data each day. It pulls in 2.7 billion like actions and around 300 million photos a day as well. Similarly, Twitter has millions of users that follow each other and form a large and complex network. We plan to retrieve data from Twitter in real time, tracking a select few hashtags (to be determined), and performing natural language processing, sentiment analysis, and other data analytics in order to draw meaning from the selected hashtags.

### (b) Parallelization

We plan to retrieve the data in parallel as not only are we doing this live, but there are also a constantly growing number of tweets, providing a pool of data large enough to make parallelization a good choice. We also plan to process the data in parallel. We may also parallelize the frontend javascript of the application, which while prone to many errors due to the scarce number of web applications that use parallelization and the sequential nature of javascript, should be useful for the enormous numbers of data we plan to display.

### (c) Implementation

We plan to use python to implement the algorithms. This is due to the ease at which python provides accessing the Twitter API due to the numerous available modules and also the compatibility of python with the web framework, Flask, which we plan to use to display the live data. For multicore processing in python and javascript, we plan to use the multithreading library and Parallel.js libraries respectively. Rust was also considered, but due to its rather young age, there has not been much documentation on it and it is not completely stable.

## References

[1]