



Tom Wallace

MSc in Big Data – Internship with Barrachd

Background and purpose

Twitter data contains a lot of networking information in addition to textual (content of tweets) and metric information (number of retweets). Not utilizing this information is a missed opportunity, but processing it requires some special techniques and tools.

The aim of this project was to build a Twitter network analysis module which can take a corpus of input tweets and automatically produce a ranking of prominent users based on the network formed by mentions, along with a visualisation.

Challenges

- Constructing and managing network data
- Ranking many users with the same score
- Showing only validated/public users
- Automated graphing
- Integrating with existing software/front-end
- Duplicates, retweets, edge weighting, etc, etc...

Solution & results

The user inputs a handle and topic that they are interested in and Python is used to collect data from the Twitter API and construct a network. Centrality scores are then used to identify how prominent that user is in the network. This is displayed as a percentile rank along with where some public handles ranked in the same network. A network visualisation is then displayed in an interactive window with the input user in orange, public users in light blue (and labelled) and private users in dark blue.

An additional module then performs a sentiment analysis on the whole network and just for tweets which mention the given user. This determines the overall mood of conversations in the network and tweets specifically mentioning the given user. This could be useful for a company tracking a social media campaign.

Front-end mock-up

