Weekly Progress Report
Detecting Conversations Going Viral

Matt Chapman

Week 16, 17/04 - 21/04

1 General Notes

- This week mostly worked on some Buzzcapture sprint stories.
- Initial data export carried out
  - Export of all Dutch twitter data nets $\approx 31M$ tweets
  - Individual queries for large businesses with annotatable spikes result in a total volume of $\approx 1.5k$ to $\approx 25k$ tweets.
  - Pandas seems fine processing large (10Gb+) files, when utilising chunking
- Experimented with data processing and representation on large CSV files using resources from NYC Open Data API.
- Added a couple more papers to the Mendeley collection
- Industry supervisor now away for rest of week.

2 Following Week

- Follow up with supervisor, get feedback on how to best approach testing algorithms
- Refine research questions, expected hypothesis
- Begin implementation of change detection algorithms
- Finish creating scaffold for simulating incoming data