

A2

Kevin R. Clemmons

February 2015

# 1 Introduction

I am absolutely embarrassed of this submission. By the time I got around to starting to do the carbon dating and graphs, I been battling a cold and I ran out of time.

## 2 Twitter Links

Retrieval of links from twitter was acquired through a script which produces a live stream of twitter data in json format on the console. To store the data the script `extract_twitter_links.py`, is run stream is piped into the file.

Processing of the json file is done in the script `linkExtractor.py` the content displayed in the actual text area of the tweet is stored via a key known as `text`. All tweets which have this are filtered through.

A regular expression is then used get figure out if a link has an html link. The techniques in doing this came from the website:

- <http://adilmoujahid.com/posts/2014/07/twitter-analytics/>

Once all links have been extracted, they are then processed to determine their redirected links and whether or not the redirected links have been encountered before. If a redirected link has been encountered, it is simply ignored. This process has helped me achieve a collection of 1000 links.

Afterwards, timemaps are retrieved and downloaded using the memento-proxy as described in the assignment description. The name of each time-map is stored in a dictionary which maps the links to their time-map file name.

Processing of the time map was performed using a class which was posted in the discussion board via the following link:

- <https://gist.github.com/jocII/c3f4a09b28fc8d590ce8>