Cover Page

PACC Project

Task 1 and Task 5

Using the SORTIF function in Excel, we created 1 hour intervals in which we counted the number of tweets, and then plotting the count values against the start of the time interval. For task 5, we filtered the retweets by using the Filter feature of Excel, and filtered out all the tweets with a value of 1 in the “Is\_Retweet” column. There seems to be no dramatic difference between the number normal tweets and number of non-retweet tweets at any point.

Task 2

There is a significant peak at 1:00 PM 10/12/2013 and a slightly smaller secondary peak at 6:00 AM 10/13/2013. There are two significant valleys at 12:00 AM 10/12/2013 and 6:00 AM 10/13/2013. Main blackouts seem to occur at 12:00 AM 10/12/2013 and 12:00 AM 10/14/2013. The cyclone made land fall at around 4:00 PM 10/11/2013, so the lack of tweets in the beginning could be a result of infrastructure interference caused by the cyclone. The peak afterwards possibly is a result of the first reports of the cyclone’s collision, and distribution of other media, prompting Twitter users to share information quickly. The general decline of the graph follows the dissipation of the cyclone, as the cyclone disappears fully on 10/14/2013, and very few tweets appear after that date.

Task 3

The frequency of “#cyclonephalin” outstrips the frequency of any other hashtag. These hashtags are mostly referring to the locations where the cyclone hit, the cyclone itself, or the newsworthiness of the cyclone. The vast majority of the tweets contained hashtags of “#cyclonephailin” by itself. It is interesting to note that most of the other tweets not containing “#cyclonephailin” combined the usage of a number the above “popular” hashtags.

Task 4

Task 6

Once more, we used the Excel Filter function, and merely deselected the appearance of all tweets with blank values for geotag information. Microsoft Excel’s Filter function quickly and efficiently provided us a new dataset of tweets only present with coordinate information.

Task 7

The removal of retweets does not seem to have introduced much sampling bias as indicated by the graph above. The removal of geotags could focus the sample to contain more tweets from more developed countries, and more serious Twitter users. Geotag data comes with smartphones, which there is a greater proliferation of in developed countries. Serious Twitter users use phones to check Twitter and post tweets, rather than casual users who could check Twitter only through computer usage, which consistently did not provide geotag data.

Task 8



We wrote a CSV file reader in Python that would read the geotag coordinates, and then plot a point onto Google Maps.

Task 9

Task 10