S3715555 – Vikas Virani – Assignment 2 – Self Assessment

Our Analysis started with exploring the twitter data & formulating a problem. We then decided to go with "Global Trends for Climate change". During the analysis, we faced many hurdles, which were resolved as we further investigated the problem. First things was to gather enough tweets which can represent the general trends about and event & can give meaningful insights. We downloaded significant tweets on **climatechange** hashtags to gather enough data. As we are doing network analysis, the next major challenge was to identify 10 actively working influencer on climate change which we can use to analyze how they form a network & interact with each other.

Next challenge for me was to properly pre-process all the raw tweets. In order to have an accurate insights, it is necessary to have a data that is pre-processed thoroughly. I've processed tweets to remove punctuations, URL & user mention filters and done stemming to normalize words. Also maintained multiple lists for further analysis while processing the data like list of tokens, hashtags, location information, date of tweets etc.

Designing an initial approach from which we can further analyze our work was a task to be think of first. There were many approaches from which we can base our further analysis. We then tried to identify geographical spread of climate awareness & what is an overall sentiment of peoples across globe for the climate changes happening.

In our Analysis phase of project, we've identified events climate prevent protest, 'Burrard Bridge' and 'Carbon footprint', identified trend detection using time based frequency ratio (words like 'canada', 'footprint', 'burrard', 'traffic' etc.) to find out the trends at particular point of time; sentiments of public about protest events. To learn about most discussed issues, topic modelling was used.

Major task was to identify geographical spread of one of the active groups in climate change, Extinction Rebellion. We've used node4j initially & Gephie to visualize a network. Given the current resources for computation, we couldn't do it for all the active groups as it took significant processing power to generate graph in node4j.

For the future plans that can be improved, if there are enough resources, a more detailed relationship can be visualized in graph using Gephie for each of the groups and how they form a network & influence people across the globe. Further modifications can be done on the way of fetching data (for example specific region based queries to identify local trends & activity level on Climate changes etc.).

The overall analysis was a great learning opportunity for all of us with many new things to come across & lessons learned.