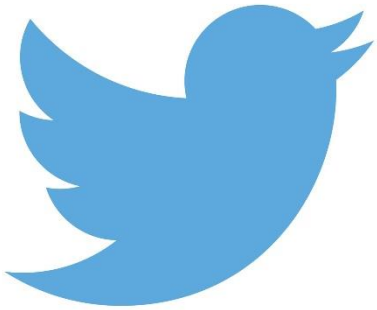


BUSA3020 Advanced Analytics Techniques

Assignment 1: Social Media Analysis

Chosen Topic #AustralianGP

Student Number: 45197083

Question 1: Report on the process, explain why you removed some observations and why you chose particular centrality measures

There are many ways to scrape tweets from Twitter; some of them involves programming while some do not. **Netyltic** is used for this assignment, as it is easy to use. In this assignment, I decided to choose to scrape data from **#AustralianGP**, as it is one of my favourite annual events every year. The 2020 Australian Grand Prix is a Formula 1 motor race that is going to take place on the 15th March 2020, in Melbourne, Australia.

Once we entered the query in **Netyltic**, it will automatically generate both a CSV file and both name (who mentions whom) and chain (who replies to whom) network. The network files can open in **Gephi** - an open-sourced network analysis and visualization software.

Once we open the graph to **Gephi**, the first step is to determine the appropriate layout as the initial graph is messy and hard to read. I decided to use ***Force Atlas***, as it is an example of a 'Force-based algorithm'. The principal of such algorithm states that the linked nodes will attract (group) together and the non-linked nodes will push further away.

Some observations are removed as those nodes did not mention any users or did not receive any mentions and RTs (Retweets). These nodes are insignificant and I have used a filter called 'Degree Range' is used to filter out the ones that have a degree of 1. In addition to filtering, centrality measures are also important as they help us to detect the most important vertices within a network. In this instance, I used network diameter as it will calculate three of the measures. The motivation, definition and the results of such measures will be explained in the following table.

Table 1: The motivation, definition and the results of centrality measures.

Centrality Name	Definition	The motivation for using such a measure.	Result + Interpretation (*Before Filter)*	Result + Interpretation (*After Filter)*
1) Network Diameter (*Undirected)*	It is the shortest distance between the two distinct nodes within a network.	The motivation of using such measure is to find out the size of the network.	11.0 – It seems that it is network a big network.	9.0 – After the filter has applied, the diameter of the network has reduced. Nonetheless, the network is still considered as a large.
2) Average path length.	It is defined as the average number of steps along the shortest paths of all possible network.	It is a useful measure as it can be used to measure the efficiency information within a network.	4.020 (Correct to 3 decimal places) – *Before Filter* 3.227 (Correct to 3 decimal places – *After Filter* - Based on research. The average path length of a social network tends to be short as it may imply that we are living in the concept of a small world where everyone is connected to everyone else through a very short path.	

Question 2: Interpret your network based only on the information in the Social Network analysis.

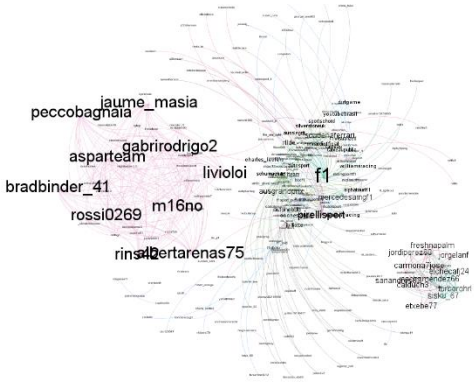
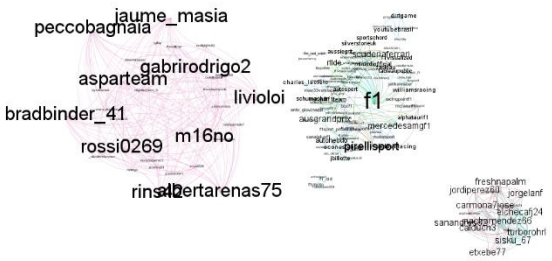
Table 2: Before Filter Social Network Analysis	
Social Network	Analysis
	There appear to be 4 main groups within the social network. The groups are motorcycle racing (MotoGP), Formula 1 (F1), news report and fandom accounts and motorsports photography-related accounts. Within the formula 1 themed group, it can be divided into four of the sub-categories. 1. Formula 1 teams such as: Mercedes AMG Petronas F1 Team, Scuderia Ferrari, Scuderia AlphaTauri and others. 2. Formula 1 Tracks such as Silverstone (a circuit in Great Britain), ausgrandprix (Australian Grand Prix) 3. Formula 1 tire supplier: Pirelli Sport 4. Formula 1 drivers such as Charles Leclerc, Max Verstappen, Antonio Giovinazzi and others. It is interesting to see that some of the Formula 1 drivers have a degree of 1 within the social media. This may imply that these drivers do not mention as much as the other drivers. These drivers include Pierre Gasly. Lewis Hamilton. Apart from Formula 1, there is also an Australian Grand Prix in MotoGP however it is hosted in Phillip Island Grand Prix Circuit. As a result, the drivers, teams and journalists are included in the social media category. One of the unique features of Twitter is to allow users to post and interact with a specific topic. In this instance, there is a group of users which they are talking about the excitement (as Australian Grand Prix is Round 1 of the Formula 1 World Championship) of a new season.

Table 3: After Filter Social Network Analysis	
Social Network	Analysis
	Filter is a tool that we can use to separate parts of the network that have different properties than others. In this instance, a filter called ***Degree Range*** is used. By using ***Degree Range***, we are able to see the network and its corresponding number of mentions within the social media network. In this instance, I decided to filter out the users which have the least mentions. This makes the graph neater and easier to identify who the main users (being mentioned) in the network.