29/10/2018

Social Media Analysis CA2

(Analysing a twitter network)

Mc Loughlin, Sean

x00130298

Contents

[Constructing the Network 2](#_Toc530997271)

[Starting Off 2](#_Toc530997272)

[My Intentions and Expectations 2](#_Toc530997273)

[Extracting the Data 3](#_Toc530997274)

[Tweet Characteristics 3](#_Toc530997275)

[Difficulties and Challenges 3](#_Toc530997276)

[Network Visualizations 4](#_Toc530997277)

[Visual Properties 4](#_Toc530997278)

[Image: 4](#_Toc530997279)

[Squares: 5](#_Toc530997280)

[Triangle: 6](#_Toc530997281)

[Layout: 7](#_Toc530997282)

[Network Interpretation 8](#_Toc530997283)

[Important Individuals 8](#_Toc530997284)

[Most Important 8](#_Toc530997285)

[Out-Degrees 8](#_Toc530997286)

[News/Media Nodes 9](#_Toc530997287)

[Analysis 10](#_Toc530997288)

# **Constructing the Network**

## **Starting Off**

The type of data that I chose to analyze for this assessment was a common word used to describe the football club, Manchester United. The word that I analyzed for this research was “mufc”. The reason why I wanted to do the twitter analysis on Manchester United is that they’re one of the most talked about club in the world and the word “mufc” would be used by many different sources such as football accounts, news accounts, sports fans account and sports accounts etc. Since the word mufc would be used by so many different sources, I thought it would be a good idea to analyze this and to show the different type of people/accounts who are talking about this topic and to show just how big they are on a global scale.

## **My Intentions and Expectations**

My intention for this research is to analyze and understand the twitter search for mufc and to also provide some visual information on the global scale of the club and that the club is talked about all around the world.

For my expectations, I’m expecting a lot of interactions with the Manchester United twitter page, I’m expecting there to be a lot of In-Degree’s for their account. I expect that that their twitter page will have the most In-degree and betweenness centrality compared to everybody else on the whole dataset.

I expect that there will be some clusters forming around news site as Manchester United is covered a lot in the media. I am hoping that this can be visually identified in the visualization dashboard.

One last thing I expect would be a lot of edges going from one cluster to another. I expect that this will happen as it is very common on twitter that sports fans follow several media sources, especially sports news, and I expect that there will be more than a few users that will be linked to multiple clusters concerning sports news on Manchester United.

## **Extracting the Data**

I extracted this data at a good time, the time I extracted this data was just after an important win for Man United and just before a huge local derby match against their biggest rivals. This was important because this way there would be a lot more of their name being mentioned in the run up to their game. I used the word “mufc” mainly because it is a common word for people to refer to the team as and nearly every media source will mention the word mufc in a tweet, whether it be a hashtag or not.

## **Tweet Characteristics**

The following calculations give an overview of the type of tweets, interaction and ways that users engaged using the MUFC word on twitter

Total Tweets: 2606

Mentions: 2606:1922 = 74%

Replies: 2606:361 = 14%

Tweets: 2606:323 = 12%

Retweets: 1922:1604 = 83% (% is the number of RT’s against the number of mentions)

Man United: 2606:197 = 7.6% (% of tweets that were directly aimed at the official page)

# Network Visualizations

## Visual Properties

There were several visual properties that I used to let the user understand that these nodes stand out. The following visual properties were used.

Image: I only used one image for this visualisation. This image is for the Manchester United twitter account. I used this as the only image as I wanted the user to understand that this image was the vocal point of the whole analyzation. By using one image, the user will fixate their eye onto the image first and will know instantly that the node they’re looking at represents Man United’s page. There are several accounts which have the same twitter picture as the official twitter account so having one photo is ideal to let users know who the real account is.

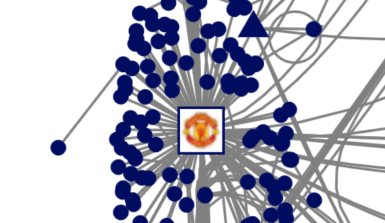


Figure : Example of the image used in the visualization

Squares: I used a custom formula to allow the user to see which account an influential account may be. Therefore, every twitter user with over 500,000 followers will have a solid square as their node in the visualization. I chose over 500,000 followers for this because there were many nodes with a high number of followers that wasn’t a reputable media source or reputable sports account. The 500,000 filters out all of these nodes and with this number, the visualization isn’t flooded with solid squares.

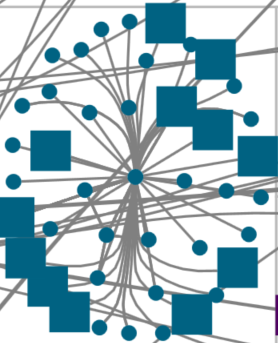


Figure : An example of where a high number of nodes with 500,000 followers were in a cluster

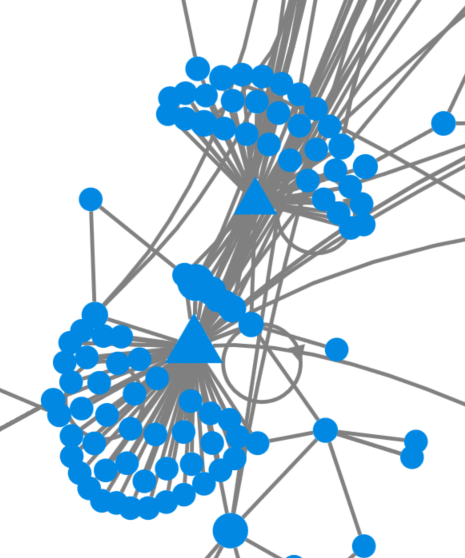
Triangle: I added a triangle shape to the visualization to inform the used that the node with that shape will be in the top 10 nodes in terms of betweenness-centrality. The only node this is void for is the Man United node as it contains an image instead. The reason I added this feature for betweenness centrality is that there were many clusters in which there was a high betweenness centrality, so I wanted to point out the 10 highest clusters for the users. 

Figure : Example of triangles being used in the visualization

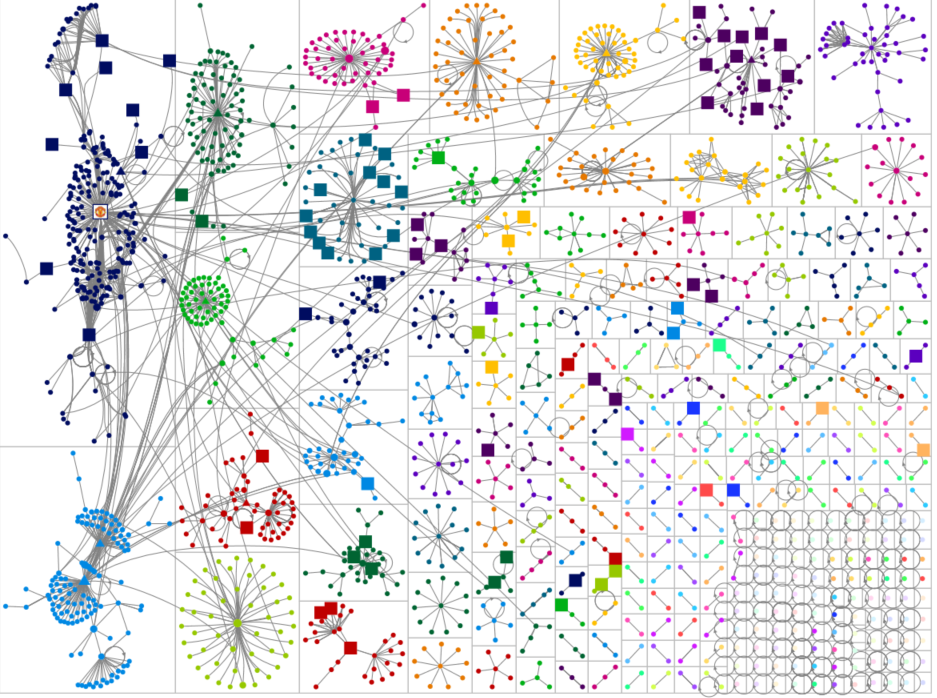
Layout:How I wanted to layout my visualisation was to first, try group up all of the nodes into different groups (clusters). By putting every node into groups, this made it easier to see who was interacting with other nodes and then it is possible to see a trend on the graph. To make the visualization appeal further to the user, I sorted the clusters into different boxes then it was possible to see the differences between all of the clusters as each cluster group was assigned a different colour. 

Figure : Layout of every cluster group

# Network Interpretation

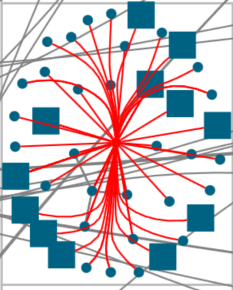
## Important Individuals

### Most Important

Unsurprisingly, one of the most important node in this visualization is the Man United twitter page. This page racked up 197 In-Degrees, the most by over 100 In-Degrees. The page also has the highest betweenness centrality compared to every other node. The page also has no out-degrees and has the highest eigenvector centrality with 0.019.

### Out-Degrees

There’s one node that’s worth noting about, this node got its own cluster group due to it having the most out-degrees with a total of 41. This node is a normal twitter user and at the time of extracting the data from twitter, this user was very active on twitter. The node doesn’t mention anything about “mufc” in their tweets, their data was extracted as “mufc” is a part of their twitter name (e.g twitter name is heyankit\_mufc).



### News/Media Nodes

Another type of node is the media/news source clusters. These clusters are very evident by looking at the data as it is a high number of in-degrees with the news node in the middle of it. There are many of these in the visualization. Another indication that these are also important users is the fact that they have a triangle as their shape which indicated that they’re in the top 10 nodes for the highest betweenness centrality.

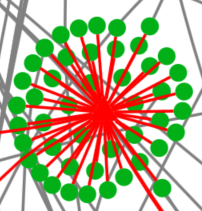
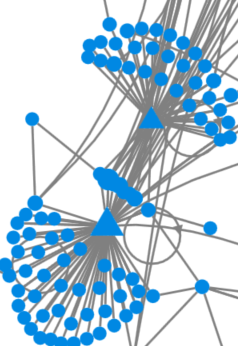
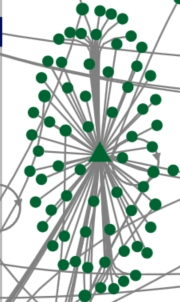


Figure : These examples show numerous clusters of news/media sources within the visualization

## Analysis

One thing I found strange after doing my analysis is the lack of out-degrees for the “influential” (500,000+ followers) nodes in this graph. Out of every node that I marked as “influential”, not one of these nodes had an out-degrees. I thought this was strange and I expected that there would certainly be at least one node with an out-degree. This shows that these so-called “influential” nodes do not communicate with their followers and they may be used for just commercial use and not for communication use.

Another stat that I found while analyzing this visualization is that there is not a lot of people communicating with each other. The majority of the nodes are not in constant communication with each other but rather taking a different approach by just retweeting. This is proved by looking at the total edges that are in the dataset which is a total of 2604, out of this figure, 1873 of these are unique edges. Out of 1873, 1604 of these are retweets. This stat may indicate that users are more interested in the content of news/media sources and not interested in communication with other people under the content.