SENTIMENT AND SOCIAL NETWORK ANALYSIS ON TWITTER DURING THE NBA 2018 FINALS

Cleveland Cavaliers vs Golden State Warriors

Amy Lee
Sonal Mendiratta
Richard (Ruiqing) He
Stella (Zhuoer) Yang
Luyao Li

TABLE OF CONTENTS

1 BACKGROUND

Chipotle gave out $1 Million worth of free burritos during the NBA Finals, which also called "Freeting" campaign, basically each time the announcers said free - Free Throw, Free Agent, Chipotle twitter account tweeted a code for free burritos. The campaign went viral and delivered successful results: delivery orders during the final 100% spiked YOY, earned 2 billion impressions yet $0 spent on sponsoring the game.

According to the research from McKinsey that millennial sports fans tune in for almost as many live sports events per week as older Gen X members and Twitter itself reports visits to its platform increase 4.1 times during live sports events to learn what's happening in real-time while traffic to other social media sites remains the same. In addition, watching these live sports events through Twitter increases more engagement and interaction with other users.
What both these stories tell us is that big sports events have a substantial impact on driving people to uniquely turn to social media platforms where real-time live streaming experiences and interactions are available.

2 OBJECTIVE

With that being said, the objective of this project is to understand the effect of the big sports games such as NBA Final on Twitter communities by analyzing users' behavior and how they are responding and interacting with every minute of the game on the Twitter platform through Sentiment Analysis and the Social Network Analysis.

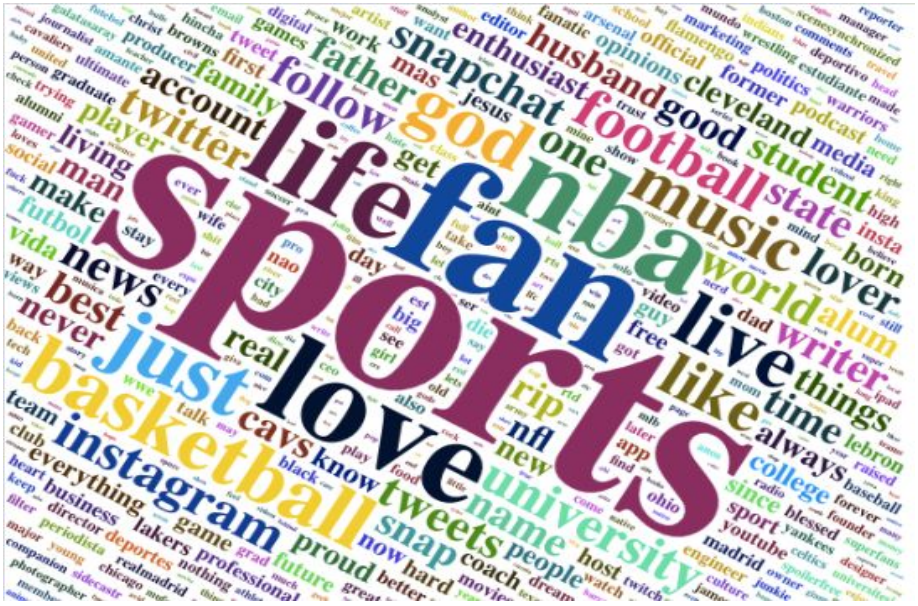
3 DATASET

The dataset is from the Kaggle notebook on "**Tweets during Cavaliers vs Warrior".** It contains Tweets captured during the 3rd game of the 2018 NBA Finals between Cleveland Cavaliers and Golden State Warriors. Tweets have been captured using Twitter's streaming API with the keyword #NBAFinals. The capture started on Thursday, June 7th, 1:13 am UTC and finished on Thursday, June 7th, 1:58 am UTC. There are two main data files, "NBA Tweets" - which include important variables for the analysis and "Location" which contain geological information such as longitude and latitude. "NBA Tweets" dataset contains the three subgroups ("Tweets", "User", "Entity") and corresponding key attributes. "Tweets" attributes are related to users' information on Twitter accounts such as Twitter id when it is created, and what they have tweeted during the game. "User" attributes are about relationships and engagement - for example, how many followers and friends the user has, how many tweets they have posted and liked. Lastly, "Entity" attributes are about extra activities such as hashtags, tweets with attached photos, and user mentions.

4 EXPLORATORY DATA ANALYSIS

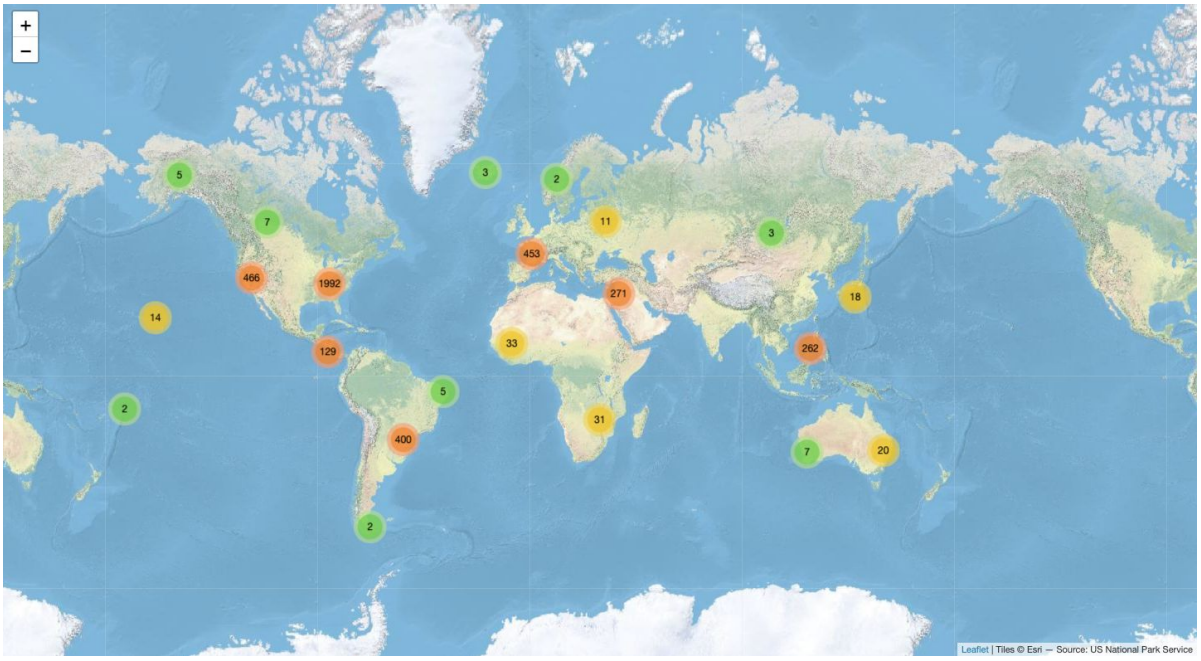
4.1 User Description

In order to see who the users are, we have extracted words from the "description" variable where people describe themselves briefly on their Twitter profiles. The captured tweets' users mostly love sports, especially in this case basketball play from both the word cloud and the bar chart below that shows the top 10 most used words on the users' profile descriptions. We do not have gender related data, but as we see "father", "husband", "man", we assume many of those users can be male. Therefore, sports fans or interested in basketball are more likely to watch the NBA and are more active on Twitter during the game.

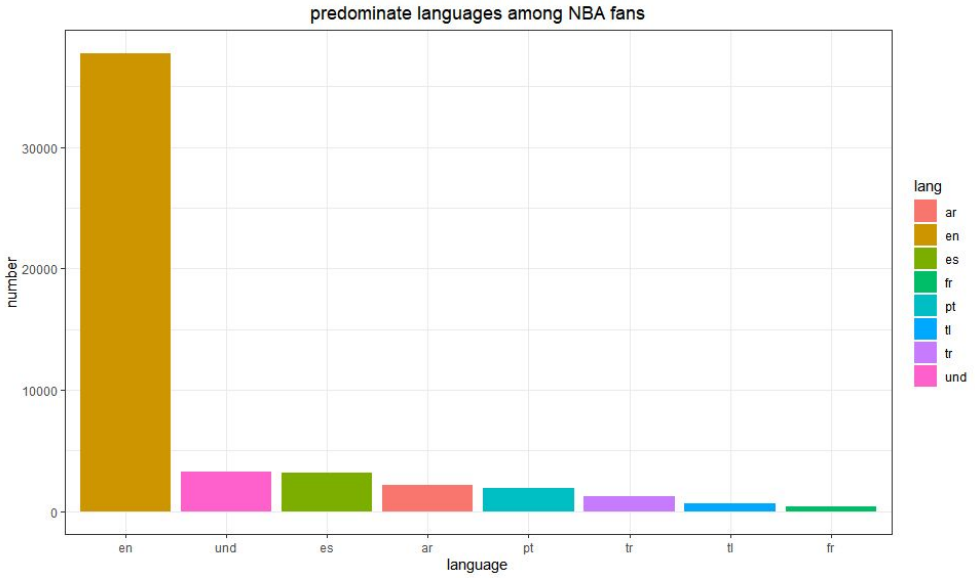


4.2 Geographic Information

Using longitude and latitude in the location dataset, the interactive map below maps out where some of the tweets (4136) were coming from during the game. We can see the majority of tweeting was happening in America followed by European and Arabic countries. This shows the NBA Final is just not about the huge game in the U.S. but is more of the game supported by fans around the world.

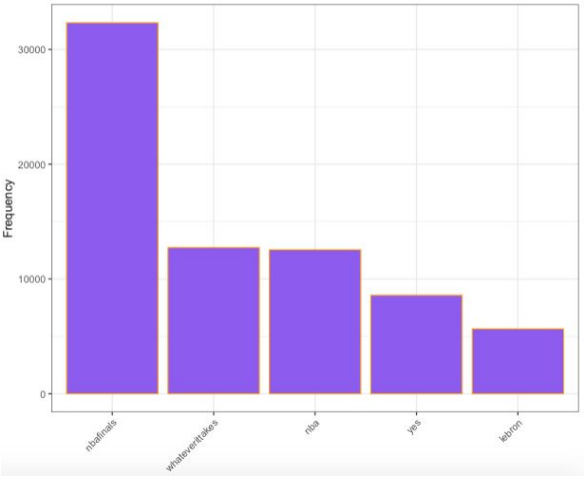


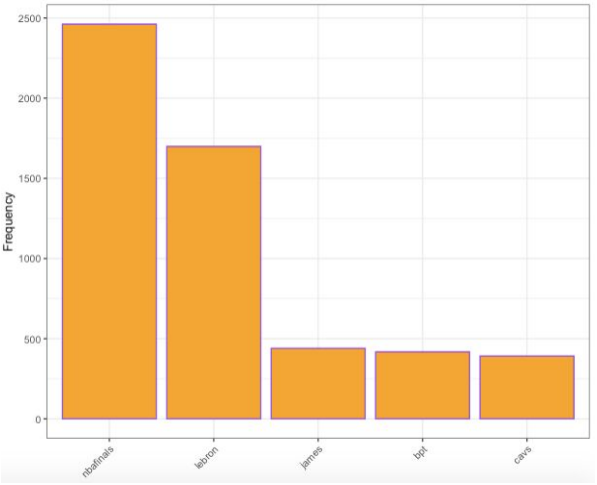
Aligned with the geographical information above, the most frequently used languages among NBA fans are “English” followed by “Spanish” and “Arabic” according to the bar plot below. *(Note “und” is undefined)*

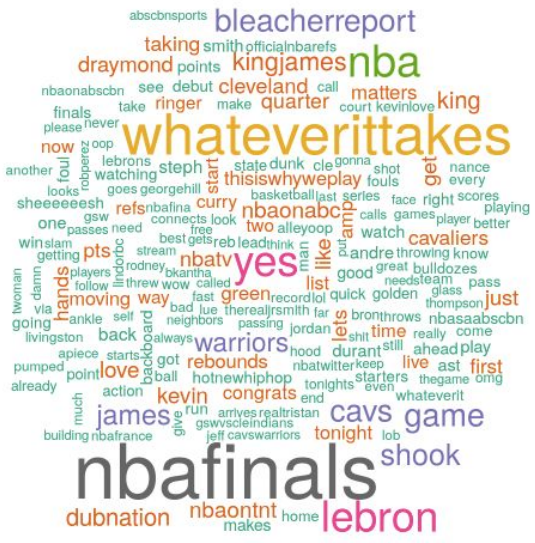


4.3 Language Usage

Since the main languages that are used among captured tweets users are “English” and “Spanish”, we wanted to take a look at what are the words they used a lot during the game as it might give us an idea of what they are interested in the most or what they are thinking during the game. It appears to be that the frequently used words are similar to one another. Both used “nbafinals” the most followed by the star player of the game, “ LeBron James”.







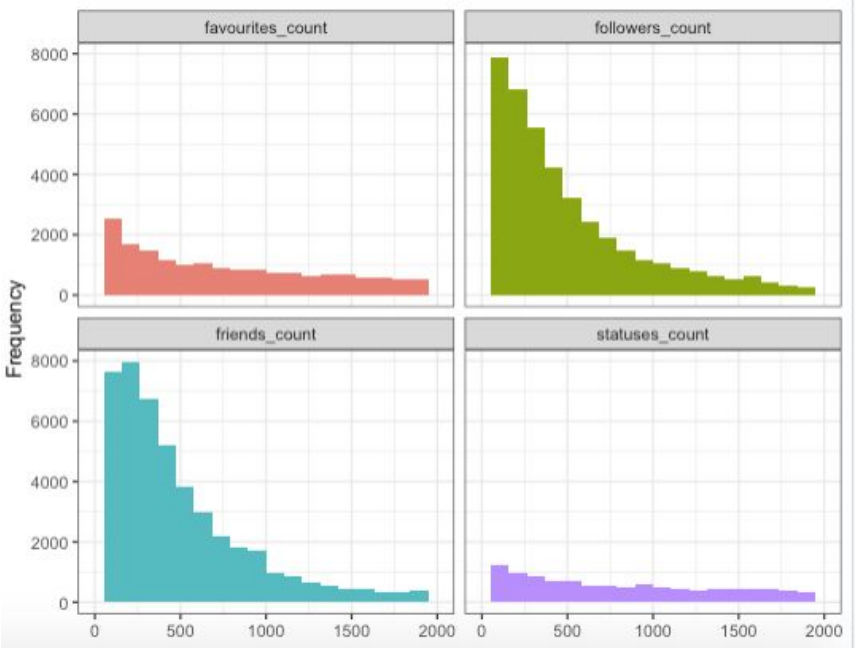
4.4 User Behavior

Understanding user behavior in terms of their activeness in the Twitter platform is important for us since it could answer whether the more they have friends the more number of tweets

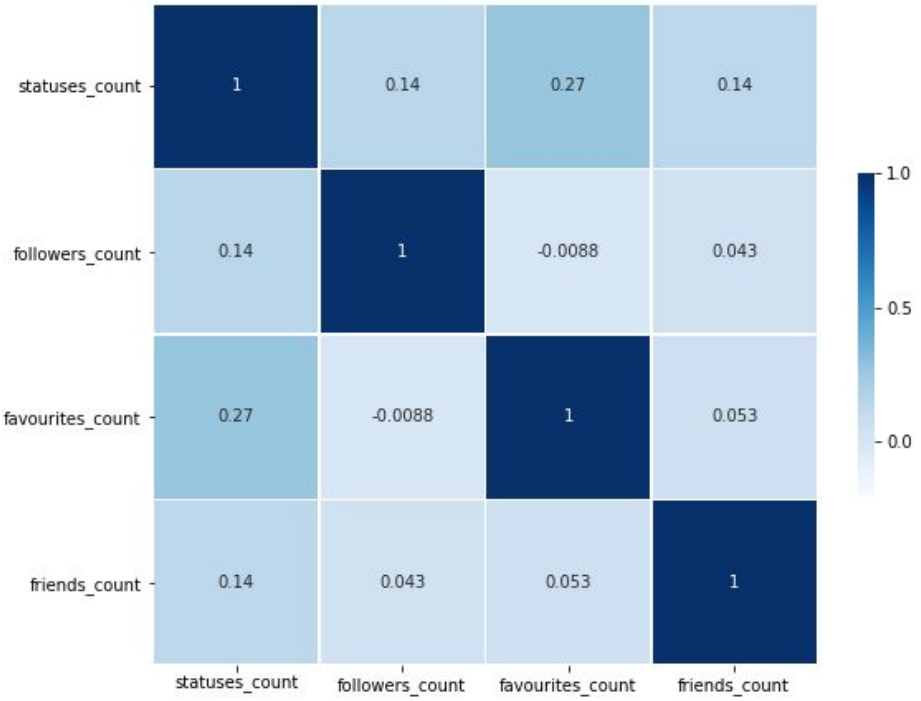
and likes ie. if there is any relationship between these covariates. Following are the four important key variables in the User attribute:

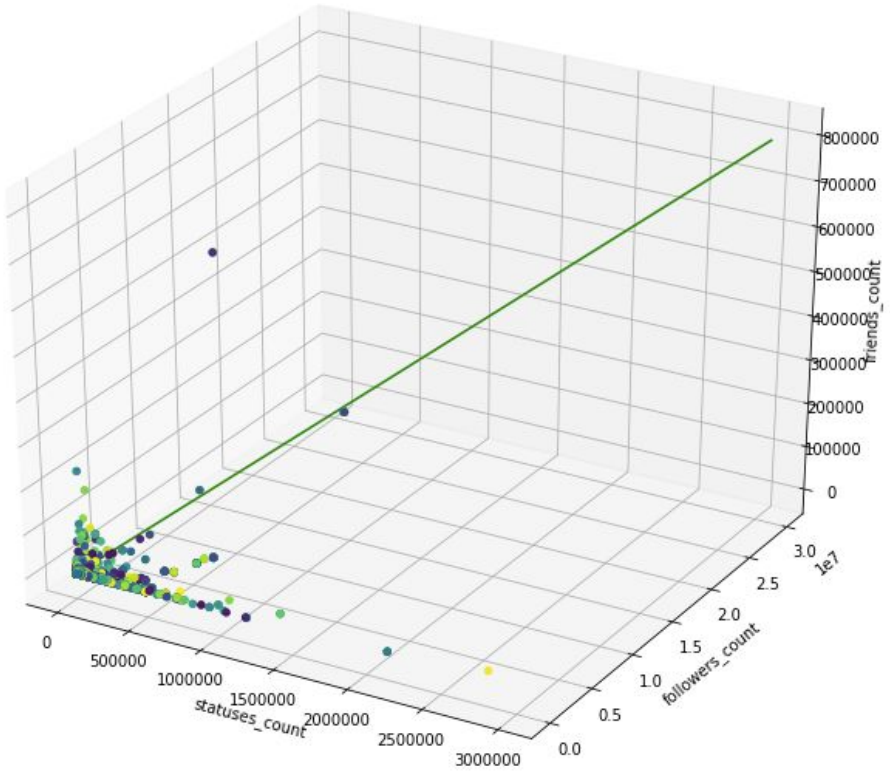
- Friends_count: The number of users this account is following.
- Followers_count: The number of followers this user currently has.
- favourites_count: The number of Tweets this user has liked in the account's lifetime.
- statuses_count: The number of Tweets (including retweets) issued by the user

Below is the histogram of each user attribute. We see the less frequency of each attribute as the number goes up. Each attribute has a long tail distribution and thus is heavily skewed.

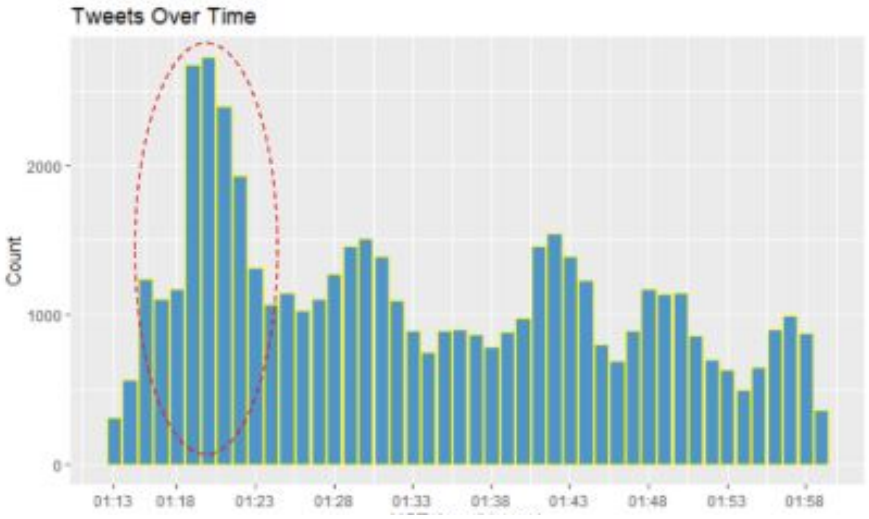


Now, we wanted to know if there are any relationships between user attributes. As mentioned above, our hypothesis was the more number of followers or friends, the more likely users have liked other tweets or the more number of tweets they post. However, there is no association between user attributes as all of the correlation coefficients are below 0.5. The highest correlation coefficient is 0.27, which is between the number of tweets user posts and the number of content users have liked.


statuses_count
followers_count
favourites_count
friends_count
1
0.14
0.27
0.14
0.14
1
-0.0088
0.043
0.27
-0.0088
1
0.053
0.14
0.043
0.053
1
1.0
0.5
0.0
statuses_count
followers_count
favourites_count
friends_count


800000
700000
600000
500000
400000
300000
200000
100000
0
friends_count
3.0
2.5
2.0
1.5
1.0
0.5
0.0
1e7
followers_count
0
500000
1000000
1500000
2000000
2500000
3000000
statuses_count

4.5 Distribution of Tweets over time during the play - Which moments are there more activity?



From the chart above, it can be seen that the number of tweets published increases sharply from 1:17 UTC. This is a key moment in the game and we will share the reason behind the spike later in the sentiment and social network analysis part.

4.6 Additional analysis after the Class Presentation

After our class presentation, as per the feedback, we wanted to understand how the scores of the teams varied throughout the game. And whether we can link the trend in the game score to the trend in the number of tweets made during the game.

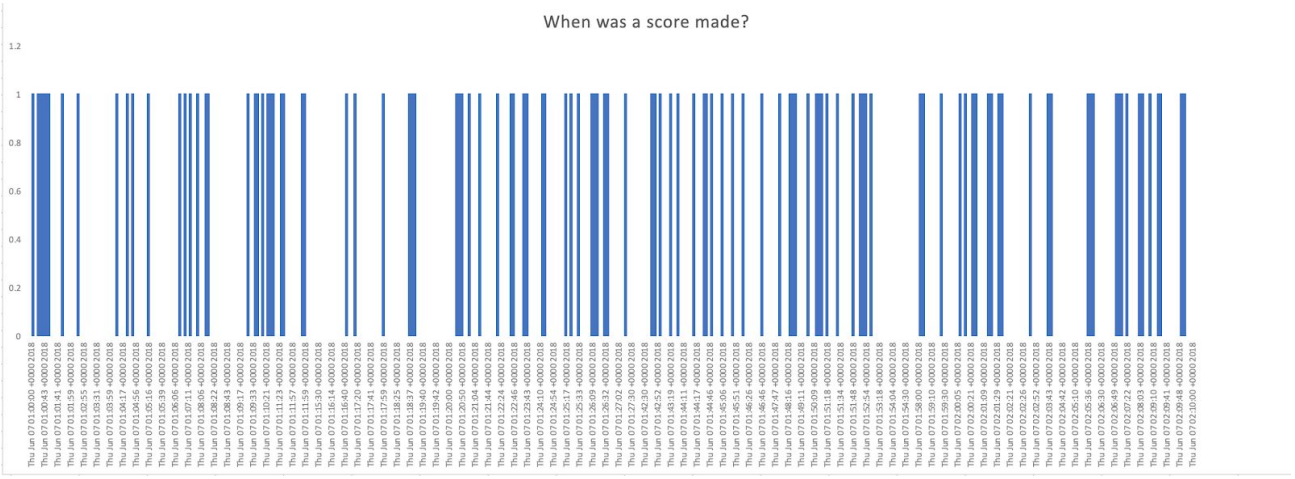
Since we did not have any such information in our data around the score during the game, we did **Webscraping** on the official ESPN website to get the minute by minute score. [https://www.espn.com/nba/playbyplay?gameId=401034615**].** The website had data on what was the score and how many minutes/seconds were left before the quarter got over. So, we had to do a lot of time manipulation to get the actual time on the UTC clock and get the play by play score details. Also, please note that for this part, our focus is on how many original tweets have been made and not whether or not they were retweeted or how many times have they been retweeted.

Below are the results that we were able to obtain after a thorough in-depth analysis.

4.6.1 When was a Score made?

We first created a binary variable to identify when a score was made in the entire game defined as, 1 - a score was made, 0 otherwise. The bars below showcase the exact time when

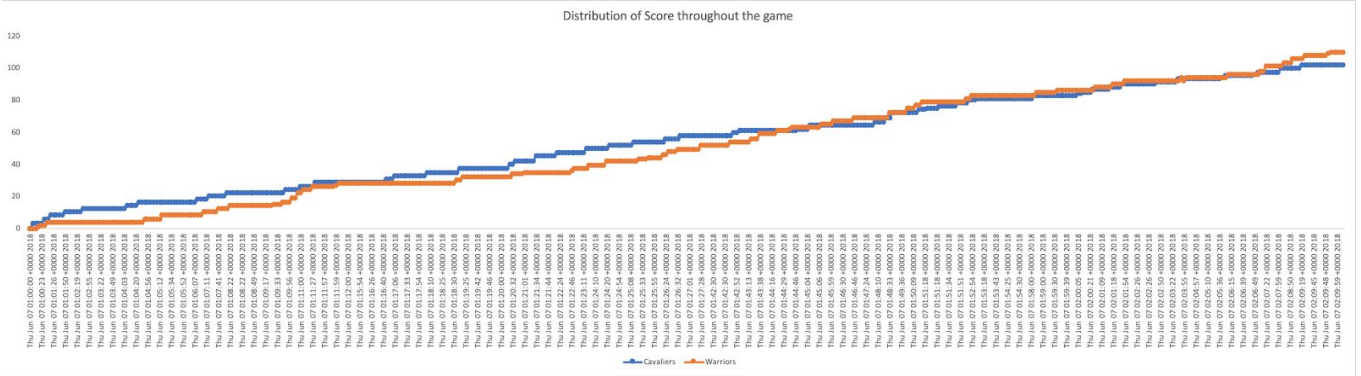
a point was made, by either team. We can also relate these timestamps with the peaks in the number of tweets made during the game in the chart in section 4.5 above. The thickness of the bar is proportional to the number of points scored in a continuous manner. For instance, if points were made successively, then the bar would be thicker.



4.6.2 Distribution of Score during the game

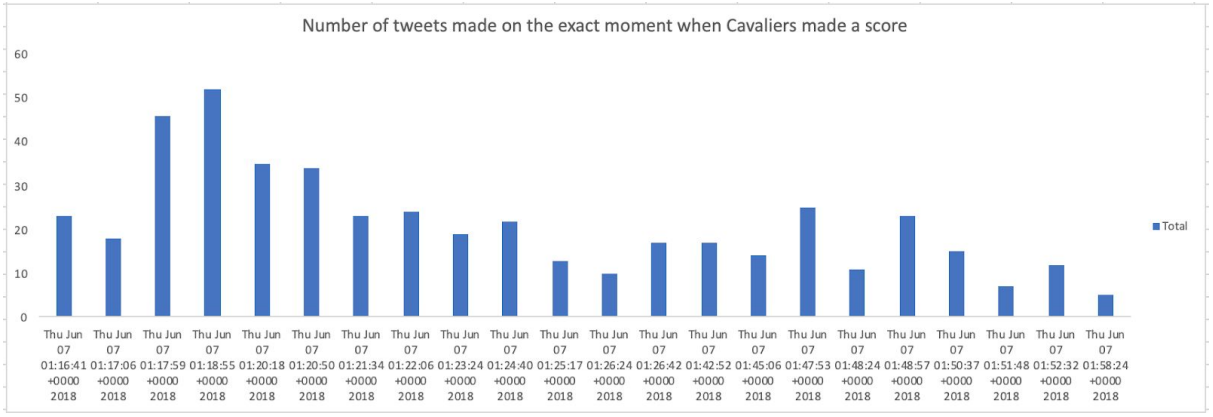
We then looked at how the scores changed for each of the teams. The graph below clearly demonstrates that it was a really interesting game where Cleaveland was leading initially because of Lebron's spectacular play but Golden State Warriors were able to catch up.It was quite a close competition!

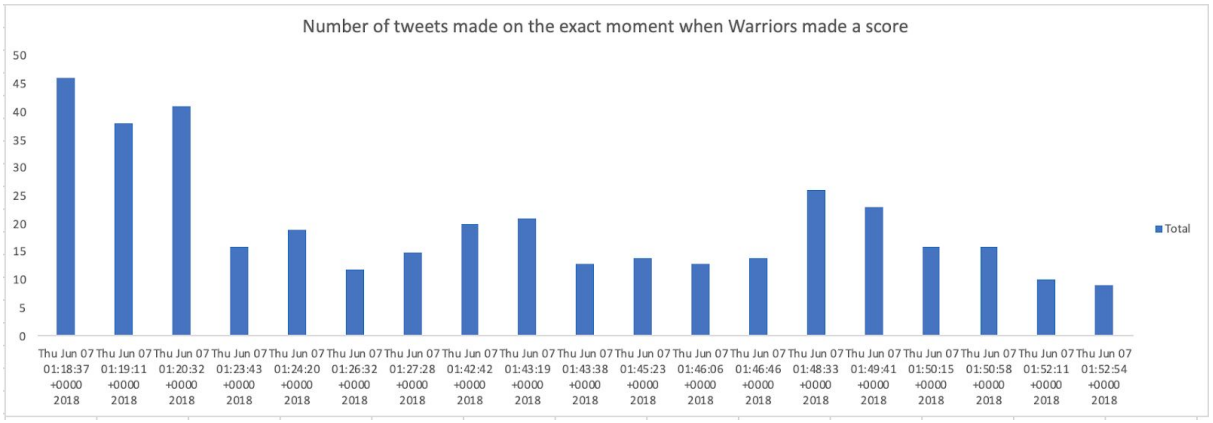
When we take a look at those moments in the score lines, we can see the spikes from the number of tweets published and the sentiments trends (1:17-20, 1:27, 1:42, 1:47; sentiments covered in section 5 later) match with the score trend lines especially when Warriors got close to lead the game or Cleveland got scores taken away from the Warriors where you may assume many emotions flowing around.



4.6.2 Tweets and Score

The ESPN data that we scraped from the web was on the entire game and every play event of the game. However, our twitter dataset captured the tweets made only during a 45 minutes window of the game(mentioned in section 3 Dataset), not necessarily aligned with the events in the game. Therefore, we had to focus only on these 45 minutes to identify the relationship between tweets and score.
We first identified the time points at which a score was made by either of the team and then found the number of tweets that were made at that **exact** same time point. The below graphs showcase these results.





These two graphs help to see how every second of the game had an impact on the tweets. For instance, at 1:17:59 UTC a score was made by Cavaliers and so the tweets were made at that time instantly and then at 1:18:37 UTC Warriors made a score and the tweets were made at that time. **So, these two graphs above show the epicentre of the flurry of tweets that followed after a few seconds.**

All in all, it was quite an impactful game, with twitter blowing up every few seconds after the above events.

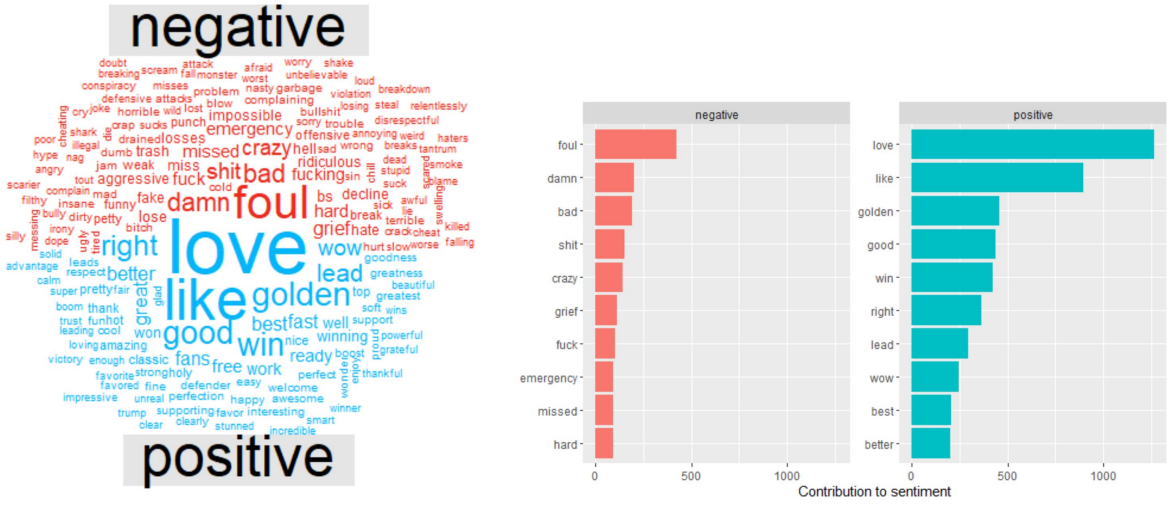
If we had the tweets for the entire game and all quarters and not just 45 minutes, we would have been able to visualize more time points when a score was made and the tweets were made.

5 SENTIMENT ANALYSIS

In this part, our main goal is to find out what kinds of sentiment categories would show in the tweets dataset, how these sentiments would differ over time, and what factors would have a significant impact on sentiment of each tweet. In order to demonstrate our findings, we split the sentiment analysis into three parts as follows.

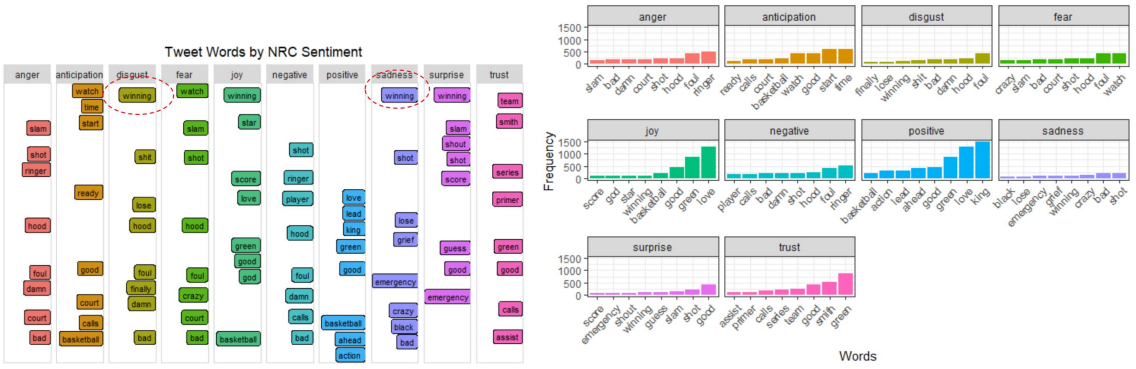
5.1 Sentiments in tweets

5.1.1 Analyze Positive and Negative Words

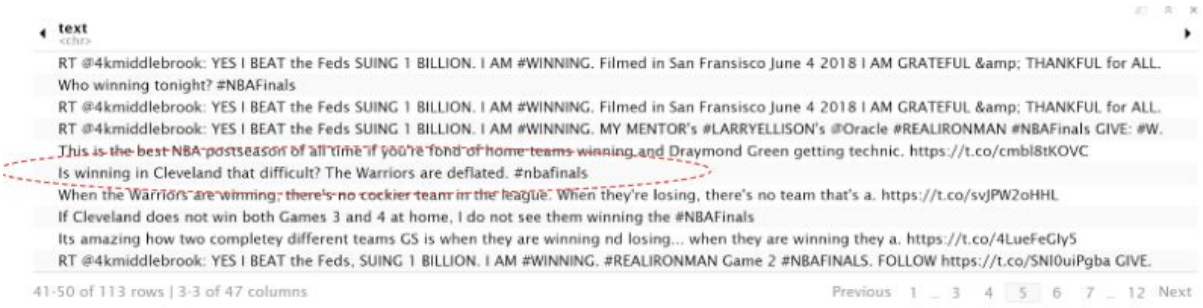


First, we made a word cloud to check how words in text could be assigned into positive and negative sentiments. The size of words in the wordcloud indicates the frequency of the usages of the words in the dataset. However, it's still vague to see how these performed in each sentiment. In order to deal with this problem, a bar plot was created to determine the top 10 words as shown on the right, and the ranking of words became clearer. Foul and Love were the most popular negative and positive words respectively.

5.1.2 More sentiments



Secondly, by using NRC LEXICON, we were able to get more sentiments from the tweets. Interestingly, we found that words like “Winning” are not only assigned to positive sentiments, but also assigned to many negative sentiments. For example, “Disgust” and Sadness. In order to understand this situation, we then took a deep look into the tweets dataset and found out many examples to explain such confusing situations.

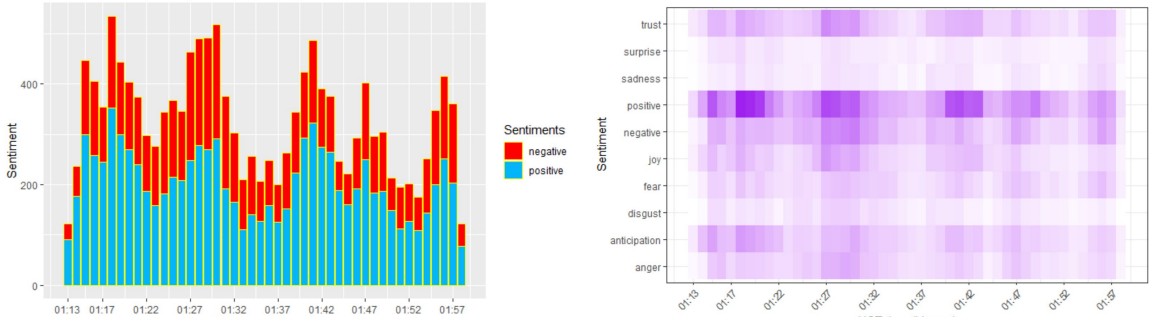


Is winning in Cleveland that difficult? The Warriors are deflated. #nbafinals

As shown in the dataset above, “Winning” in sentence “Is winning in Cleveland that difficult? The Warriors are deflated. #nbafinals” could indicate disgust and sadness.

5.2 Sentiment Over Time (UTC)

Changes in sentiments at a granular level



In this section, we want to see how sentiment would differ over time, and our findings are:

- There are 5 humps in the left chart during the time periods, which means there should be 5 sentiment climaxes during the time periods. Most of the time positive sentiment found to be more than the negative sentiment overall during the captured time. The humps also aligned with the brightness of the color trends of the graph on the right side as we found the strong color scheme is seen between the climax 1:17-1:20, 1:27, 1:42, 1:47. Positive, trust, and anticipation sentiments are dominating around those times and this could mean something positive happened during the game and evoked users' emotions. Interestingly, negative and angers are spotted as well especially on the second hump (1:27) , we assume the event happened not just joy but also brought anger and sadness to some users. We will look into what happened those times later on.
- The behaviors of positive sentiment and negative sentiment are in a similar trend – they go up and down in a similar way. This also confirms the event that happened during the spikes could be positive events for some users but could also be negative events for others.

5.3 Multinomial Logistic Regression

5.3.1 Model definition

$$\ln\left(\frac{P(sentiment = 0)}{P(sentiment = -1)}\right) = \beta_0 + \beta_1 statuses\_count + \beta_2 followers\_count + \beta_3 favourites\_count + \beta_4 friends\_count + \beta_5 word\_count$$

Since we have three categories (which are positive, negative and neutral sentiment) in our target variable, we decided to use the multinomial logistic regression model for modeling part. As shown in the formula above, the basic idea of multinomial logistic regression is to set up a baseline first and then do the comparisons between other variables and this baseline one

by one.

```
Call:
nnet::multinom(formula = sentiment ~ statuses_count + followers_count +
    favourites_count + friends_count + word_count, data = tweet_sentence_data)


[1] "z-score:"
  (Intercept) statuses_count followers_count favourites_count friends_count word_count
0 87279653958      -4.916411        1.233620         4.346435    -0.3773605 -293385477
1 56025622206      -8.173532        1.416178         4.554549    -0.3234826  -45298320


[1] "p-value:"
  (Intercept) statuses_count followers_count favourites_count friends_count word_count
0           0    8.814508e-07       0.2173445     1.383683e-05     0.7059057          0
1           0    2.220446e-16       0.1567235     5.249815e-06     0.7463293          0
```

In the regression result, p-values are very close to 0 which means the statuses_count, favourite_count and word_count statistically have a significant impact on the sentiment of each tweet. Therefore, this tells us that statuses_count, favorite_count , and word_count are meaningful variables to understand the users' changes in emotions during the game.

5.3.2 Model interpretation

```
# weights:  21 (12 variable)
initial  value 56496.136945
iter  10 value 53633.087729
iter  20 value 43833.246300
final  value 43830.079260
converged
Call:
nnet::multinom(formula = sentiment ~ statuses_count + followers_count +
    favourites_count + friends_count + word_count, data = tweet_sentence_data)

Coefficients:
  (Intercept) statuses_count followers_count favourites_count friends_count  word_count
0    3.961389  -9.977904e-07    6.571003e-08     2.608277e-06 -9.118177e-07 -0.17901804
1    1.607169  -1.695739e-06    7.316983e-08     2.688346e-06 -6.230760e-07 -0.01682784

Std. Errors:
   (Intercept) statuses_count followers_count favourites_count friends_count   word_count
0 4.538732e-11   2.029509e-07    5.326602e-08     6.000958e-07  2.416304e-06 6.101803e-10
1 2.868633e-11   2.074671e-07    5.166712e-08     5.902552e-07  1.926150e-06 3.714893e-10

Residual Deviance: 87660.16
AIC: 87684.16
```

- The variable **statuses_count** has the **negative impact** on the log odds of being neutral sentiment or being positive sentiment vs. negative sentiment which means captured tweets users who issued more the number of tweets are expected to get the negative sentiments than neutral or positive sentiments.

- The variable **favourite_count** has the **positive impact** on the log odds of being neutral sentiment or being positive sentiment vs. negative sentiment. In other words, the more the number of tweets users have liked in the account's lifetime, the more likely it is that the sentiment is going to be positive.
- The variable **word_count** has the **negative impact** on the log odds of being neutral sentiment or being positive sentiment vs. negative sentiment, which means the more words a sentence has the more likely it is that the sentiment is going to be negative than positive or neutral. Possibly what we can do in the future to extend this results is that we can see if a trend of the number of words/characters during the game and see if we can get the peak of using words matches the peak of the tweets published.

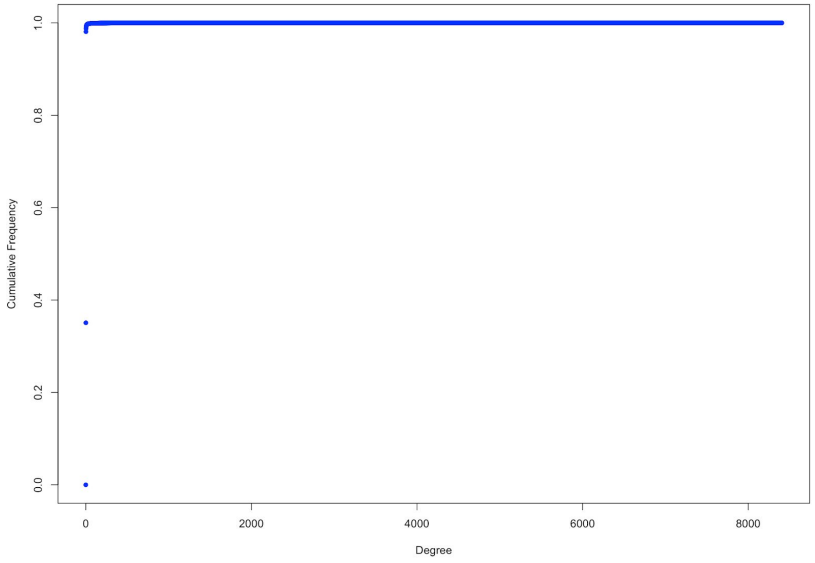
6 SOCIAL NETWORK ANALYSIS

We wanted to analyze how the tweets were impacted during the game and therefore performed Social Network Analysis on twitter data. Our data had rich information around which tweet was retweeted by which twitter account. We summarized it to create a network and defined the tweets as the nodes and all the retweets as the edges.In all, we had 53,047 tweets and 31,301 edges.

6.1 Statistical Measures

a. Degree Distribution

The degree distribution of a graph is the probability distribution of the degrees over the entire network.Our degree distribution graph indicates that degree increases at a gradually slow rate.

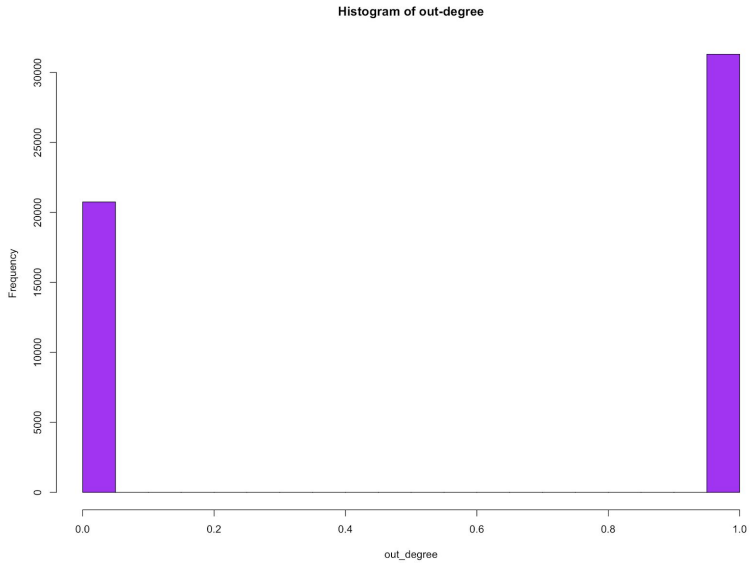


b. Edge Density

The value for Edge Density of the network is 1.155514e-05 which is very low for 2 reasons. Retweets are only outflows from Source to Target and most users are “content consumers” who only retweet, rather than creating content for others to retweet.

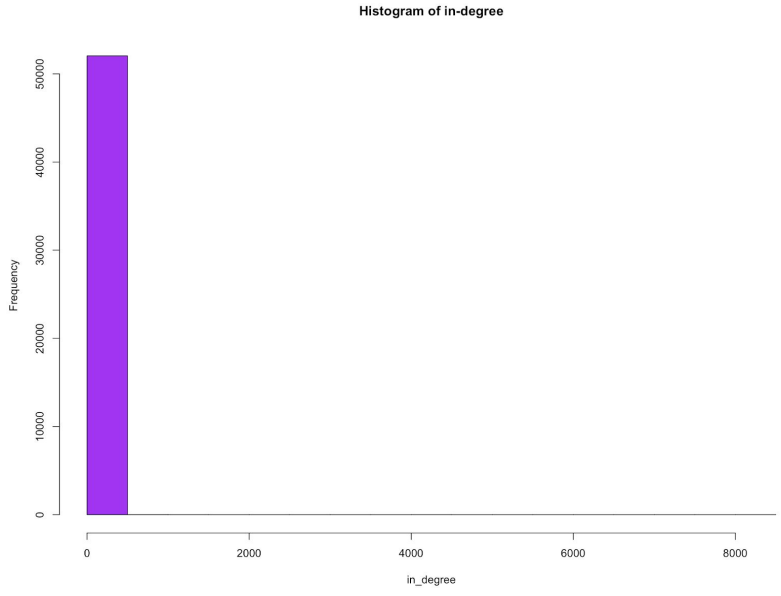
c. Out Degree Distribution

As shown by the chart below, the out degrees are either 0 or 1 because in the network the trend is that one node makes a tweet which is either retweeted or not retweeted.Out-degree is the number of connections that originate at a vertex and point outward to other vertices.



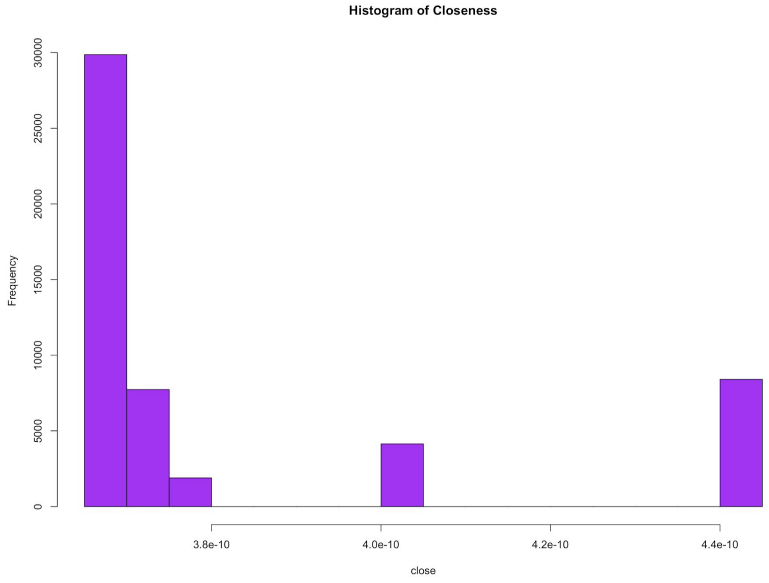
d. In Degree Distribution

In-degree is the number of connections that point inward at a vertex. In our network, the in-degree distribution is highly skewed as some tweets from main users are retweeted many times while the majority are not retweeted.



e. Closeness

Closeness is the inverse of the node's average geodesic distance to others in the network.The chart below shows that nodes are very close to each other because of the inherent definition of how an edge is defined.



f. Reciprocity

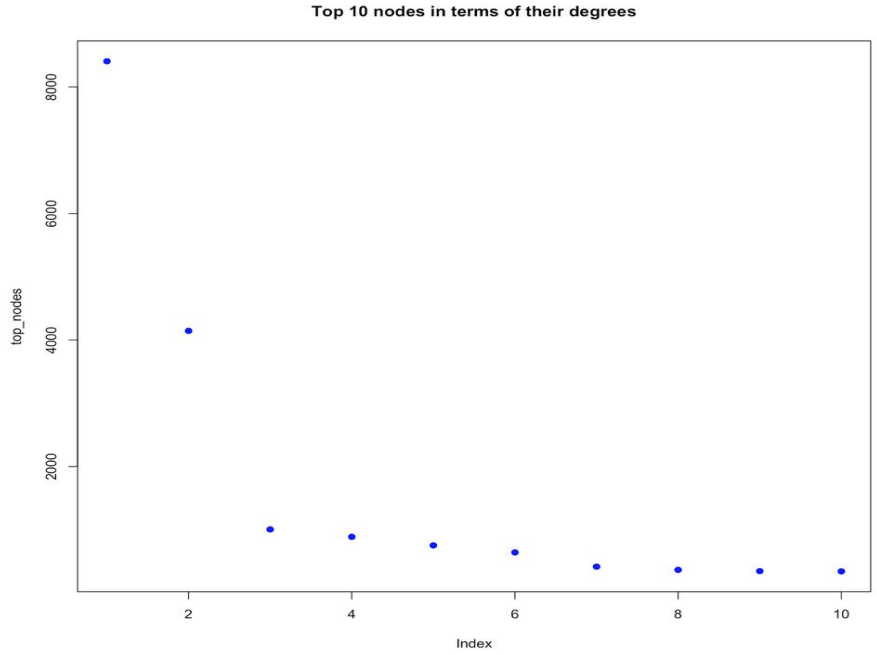
Reciprocity is the number of relations that are reciprocated such as an edge in both directions over the total number of relations in the network. In our case, this value is 0 because there are no bi directional links.

g. Degree Centrality

It shows us how many direct, ‘one-hop’ connections each node has to other nodes in the network and is useful in assessing which nodes are central with respect to spreading information and influencing others’ in their immediate neighborhood. The value of Centralization is 0.08074548 which is very high. This is further validated by our visualizations in the section 6.3.

h. Top 10 nodes in terms of degree

The following graph shows the degrees of the top 10 nodes in the network.The max degree is 8406 which means that the most popular tweet got retweeted 8406 times which was 100 times more than the second most popular tweet was retweeted around 4000 times.



The corresponding accounts from which these top 10 most popular tweets were :

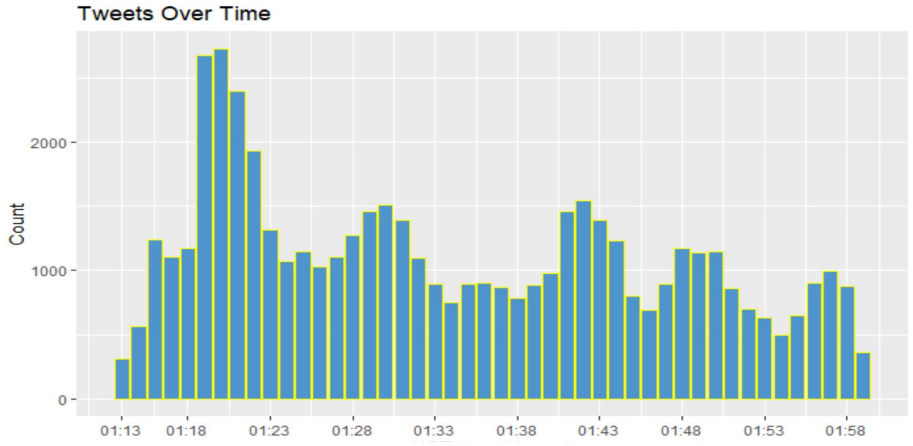
- NBA
- BleacherReport
- Arabic1_NBA
- NBAonTNT
- NBATV
- NBA
- B24PT
- warriors
- BleacherReport
- HotNewHipHop

Apart from the official NBA account and the sports website Bleacher Report, interestingly an Arabic user was in the top 3 users which is inline with our result on the geographic representation earlier.

i. Most Retweeted Tweet = Node with the highest degree

Using the URL of the tweets we were able trace back to these actual tweets and the most popular tweet was when Lebron James made a impressive dunk:

As we promised, we now can reveal the secret of the spike of the graph below that we also shared in the EDA part. This is exactly the time in the game, where we saw a huge jump in the number of tweets during the finals as depicted by the below graph, between 1:17UTC and 1:20UTC.



Which words were most popular during peak time?

Out of curiosity, we wanted to see what are the words users used the most during the peak time (1:17 am UTC - 1:20 am UTC). We first extracted tweets only published between the peak time and found 451 tweets. As we can see below in the wordcloud, the star of the night "**Lebron James**' ' is the most popular words followed by the "**point**", "**game**", and"**fan**". This shows the twitter users are on the spot and share the game experience in real-time.

j. Second Most Retweeted Tweet = Node with the second highest degree

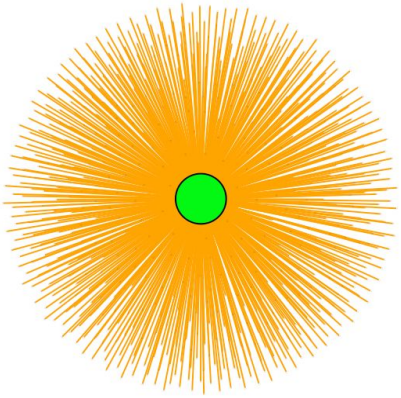
The second most retweeted tweet was a millennial slang, ‘Shook’ with the official hashtag NBAFinals.

Please note that the retweeted counts in the images above are more than those reported in our analysis because these numbers are the latest figures.

6.2 Visualization on R

We tried various methods to visualize the network on R for the top 10 nodes individually, but the resultant networks did not help in clearly depicting the relationship. This is because the main node made a tweet which was retweeted once by multiple users thereby resulting in

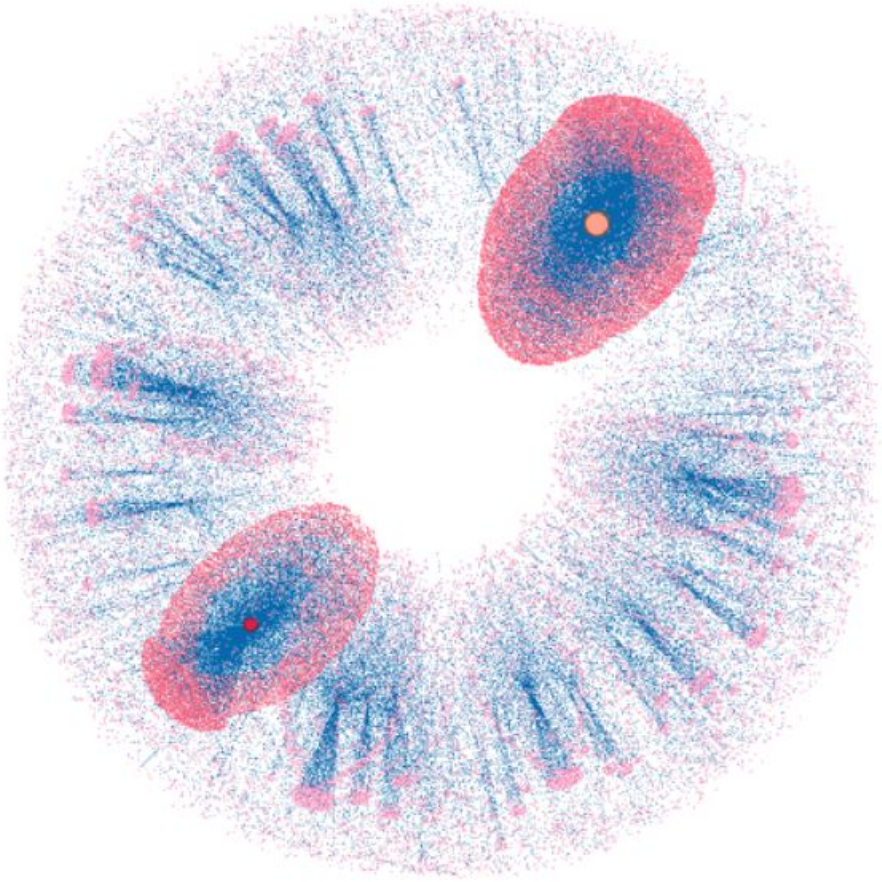
huge difference in the degrees of all the nodes in the network.For instance, the network for the highest node had 8406 nodes in all with the nodes which are part of the edges have degree equal to 1. Thus the corresponding network below was not informative.



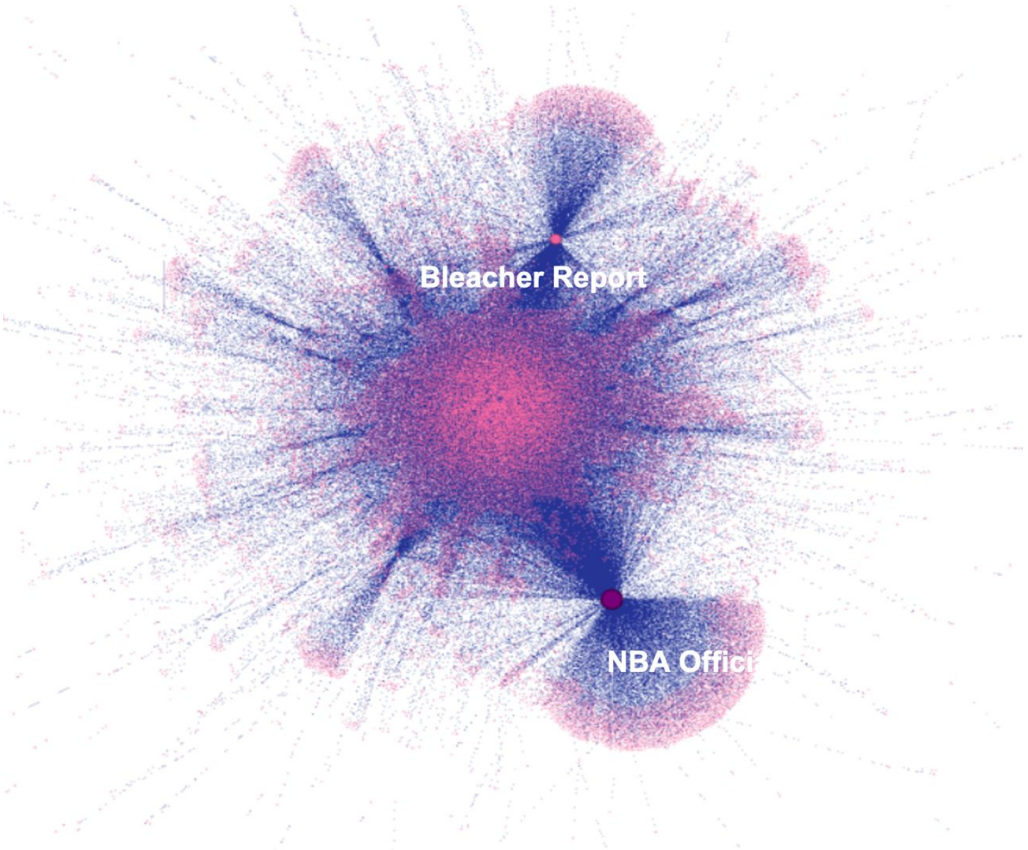
Since R was not capable of visualizing the networks because of the sheer size of the data, we decided to explore the tool Gephi. The following section covers the details of the network analysis done using it.

6.3 Visualization on GEPHI

We first cleaned the data to have only “Source” and “Target”, then we labeled each row of data by their id, which is considered “Nodes”. Different types of visualization techniques were explored. The first one is called “Frutcherman Reingold”, which weighs in the gravity of the nodes by their degrees:

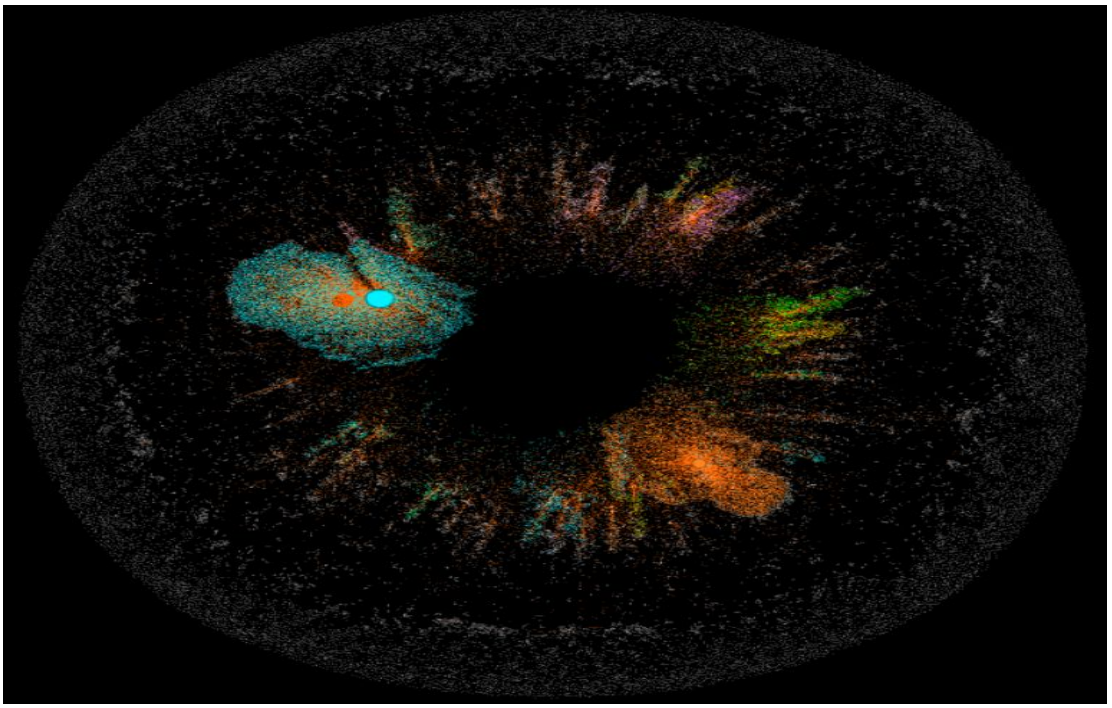


This is clear that the two nodes have the strongest impact: on the top right corner is the **NBA official account**. On the lower-left corner is the **Bleacher report's account.**
Secondly, with further inspection, we used the "Yifan Hu" layout algorithm to map the data. The Yifan Hu Multilevel layout algorithm is an algorithm that brings together the good parts of force-directed algorithms and a multilevel algorithm to reduce algorithm complexity. This is one of the algorithms that work really well with large networks.



After seeing the layout, it is interesting to us that the cluster in the center represents that there are many common fans there are for both Bleacher Reports and NBA Officials.

The last one is called “ForceAtlas”, which forces nodes to move away from each other and create clusters that share common posts/retweets. With some color coding, it is obvious to see multiple interactions within the network.



7 CONCLUSION AND KEY TAKEAWAYS

We have come to the conclusion that major sports games can have a substantial impact on the Twitter community and the interaction between users after we have analyzed the tweets during the NBA Final game through sentiment analysis as well as the social network analysis. What users are most interested in every minute during the game is reflected by the words they use in tweets and what they retweeted. We were able to confirm this by looking at the most retweeted tweets (James Lebron got a dunk shot) during the game and the words used in most during the peak time of user activity (James Lebron, game, point). Leveraging these results, marketing agencies can tap into emotions/behaviors and can start a hashtag trend or guerilla marketing campaign regarding the game especially when users are most active and twitter is blowing up. We also learned that most of the captured users are sports fans and the emotions of users during the game were quite diverse based on our sentiment analysis. With this, in the future twitter can have ads on their platform on sports fans' wildly changing emotions.

8 APPENDIX

8.1 Additional Statistical Measures of the Network

a. Eigen Centrality

Eigenvector centrality is a measure of the influence a node has on a network. If a node is pointed to by many nodes (which also have high Eigenvector centrality) then that node will have high eigenvector centrality.

```
$vector
1004531741216989191 1004531741422481409 1004531741954981888 1004531743410573312 1004531743272194048 1004531744010272768 1004531743976718336
                  0                   0                   0                   0                   0                   0                   0
1004531744081575936 1004531743666286592 1004531744291336193 1004531744337522689 1004531745222348800 1004531745088303109 1004531745658728448
                  0                   0                   0                   0                   0                   0                   0
1004531746895851520 1004531746765860864 1004531747340550144 1004531747663511552 1004531748284313608 1004531748481445889 1004531748691009536
                  0                   0                   0                   0                   0                   0                   0
1004531748552626177 1004531751358795776 1004531752537415683 1004531752864571394 1004531753703428096 1004531754013790209 1004531754366103552
                  0                   0                   0                   0                   0                   0                   0
1004531754533900288 1004531755435659265 1004531756408680450 1004531757612486656 1004531757973110785 1004531753887961096 1004531758904332288
                  0                   0                   0                   0                   0                   0                   0
1004531759390904320 1004531759604617216 1004531760816951301 1004531761517342721 1004531761899024390 1004531762050097157 1004531762049921024
                  0                   0                   0                   0                   0                   0                   0
1004531762364633088 1004531763056693249 1004531765590089730 1004531765741084674 1004531762700193792 1004531766064041984 1004531767615860741
                  0                   0                   0                   0                   0                   0                   0
1004531767838199808 1004531768261791745 1004531768714641408 1004531768949604354 1004531769167659008 1004531769448726528 1004531769683664896
```

b. Betweenness

It measures the number of times a node lies on the shortest path between other nodes.

```
> between
1004531741216989191 1004531741422481409 1004531741954981888 1004531743410573312 1004531743272194048 1004531744010272768 1004531743976718336
                  0                   0                   0                   0                   0                   0                   0
1004531744081575936 1004531743666286592 1004531744291336193 1004531744337522689 1004531745222348800 1004531745088303109 1004531745658728448
                  0                   0                   0                   0                   0                   0                   0
1004531746895851520 1004531746765860864 1004531747340550144 1004531747663511552 1004531748284313608 1004531748481445889 1004531748691009536
                  0                   0                   0                   0                   0                   0                   0
1004531748552626177 1004531751358795776 1004531752537415683 1004531752864571394 1004531753703428096 1004531754013790209 1004531754366103552
                  0                   0                   0                   0                   0                   0                   0
1004531754533900288 1004531755435659265 1004531756408680450 1004531757612486656 1004531757973110785 1004531753887961096 1004531758904332288
                  0                   0                   0                   0                   0                   0                   0
1004531759390904320 1004531759604617216 1004531760816951301 1004531761517342721 1004531761899024390 1004531762050097157 1004531762049921024
                  0                   0                   0                   0                   0                   0                   0
1004531762364633088 1004531763056693249 1004531765590089730 1004531765741084674 1004531762700193792 1004531766064041984 1004531767615860741
                  0                   0                   0                   0                   0                   0                   0
1004531767838199808 1004531768261791745 1004531768714641408 1004531768949604354 1004531769167659008 1004531769448726528 1004531769683664896
                  0                   0                   0                   0                   0                   0                   0
```

c. Hub Score

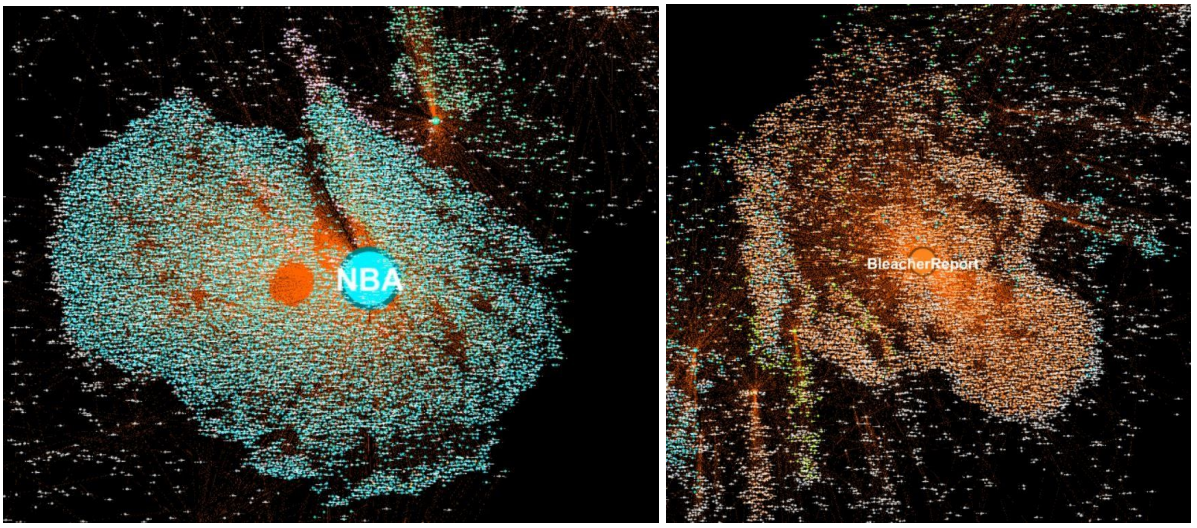
The hub score (hs) measures the out-going links for a given node, whereas the authority score (as) measures the in-coming links for a given node. Typically, a good hub (the node with high hub score) points to good authorities and a good authority (the node with high authority score) is pointed to by a good hub.

```
> hs
1004531741216989191 1004531741422481409 1004531741954981888 1004531743410573312 1004531743272194048 1004531744010272768 1004531743976718336
       2.671489e-16        2.112340e-17        2.112340e-17        2.671489e-16        1.491064e-16        4.870808e-16        2.112340e-17
1004531744081575936 1004531743666286592 1004531744291336193 1004531744337522689 1004531745222348800 1004531745088303109 1004531745658728448
       2.112340e-17        2.112340e-17        1.975660e-16        8.076596e-17        2.112340e-17        8.076596e-17        2.112340e-17
1004531746895851520 1004531746765860864 1004531747340550144 1004531747663511552 1004531748284313608 1004531748481445889 1004531748691009536
       2.112340e-17        4.224681e-17        2.112340e-17        2.671489e-16        2.112340e-17        3.267915e-16        2.671489e-16
1004531748552626177 1004531751358795776 1004531752537415683 1004531752864571394 1004531753703428096 1004531754013790209 1004531754366103552
       2.671489e-16        1.354383e-16        8.076596e-17        2.112340e-17        2.671489e-16        4.870808e-16        2.112340e-17
1004531754533900288 1004531755435659265 1004531756408680450 1004531757612486656 1004531757973110785 1004531753887961096 1004531758904332288
       2.112340e-17        2.112340e-17        8.076596e-17        2.385702e-16        8.076596e-17        2.112340e-17        2.112340e-17
1004531759390904320 1004531759604617216 1004531760816951301 1004531761517342721 1004531761899024390 1004531762050097157 1004531762049921024
       2.112340e-17        1.491064e-16        4.224681e-17        2.112340e-17        8.076596e-17        2.112340e-17        5.467234e-17
1004531762364633088 1004531763056693249 1004531765590089730 1004531765741084674 1004531762700193792 1004531766064041984 1004531767615860741
       1.540766e-16        2.112340e-17        3.267915e-16        2.112340e-17        2.112340e-17        2.112340e-17        2.112340e-17
1004531767838199808 1004531768261791745 1004531768714641408 1004531768949604354 1004531769167659008 1004531769448726528 1004531769683664896
       2.112340e-17        2.671489e-16        2.112340e-17        2.112340e-17        4.758979e-16        2.112340e-17        2.485106e-16
```

d. Authority Score

1004531741216989191	1004531741422481409	1004531741954981888	1004531743410573312	1004531743272194048	1004531744010272768	1004531743976718336
5.365446e-17	5.365446e-17	5.365446e-17	5.365446e-17	5.365446e-17	5.365446e-17	5.365446e-17
1004531744081575936	1004531743666286592	1004531744291336193	1004531744337522689	1004531745222348800	1004531745088303109	1004531745658728448
5.365446e-17	5.365446e-17	5.365446e-17	5.365446e-17	5.365446e-17	5.365446e-17	5.365446e-17
1004531746895851520	1004531746765860864	1004531747340550144	1004531747663511552	1004531748284313608	1004531748481445889	1004531748691009536
5.365446e-17	1.072818e-16	5.365446e-17	5.365446e-17	5.365446e-17	5.365446e-17	5.365446e-17
1004531748552626177	1004531751358795776	1004531752537415683	1004531752864571394	1004531753703428096	1004531754013790209	1004531754366103552
5.365446e-17	5.365446e-17	5.365446e-17	5.365446e-17	5.365446e-17	5.365446e-17	5.364090e-17
1004531754533900288	1004531755435659265	1004531756408680450	1004531757612486656	1004531757973110785	1004531753887961096	1004531758904332288
5.365446e-17	5.365446e-17	5.365446e-17	5.365446e-17	5.365446e-17	5.365446e-17	5.365446e-17
1004531759390904320	1004531759604617216	1004531760816951301	1004531761517342721	1004531761899024390	1004531762050097157	1004531762049921024
5.365446e-17	5.365446e-17	1.072818e-16	5.365446e-17	5.365446e-17	5.365446e-17	1.608956e-16
1004531762364633088	1004531763056693249	1004531765590089730	1004531765741084674	1004531762700193792	1004531766064041984	1004531767615860741
5.365446e-17	5.365446e-17	5.365446e-17	5.365446e-17	5.365446e-17	5.365446e-17	5.365446e-17
1004531767838199808	1004531768261791745	1004531768714641408	1004531768949604354	1004531769167659008	1004531769448726528	1004531769683664896
5.364090e-17	5.365446e-17	5.365446e-17	5.365446e-17	5.365446e-17	5.365446e-17	5.365446e-17

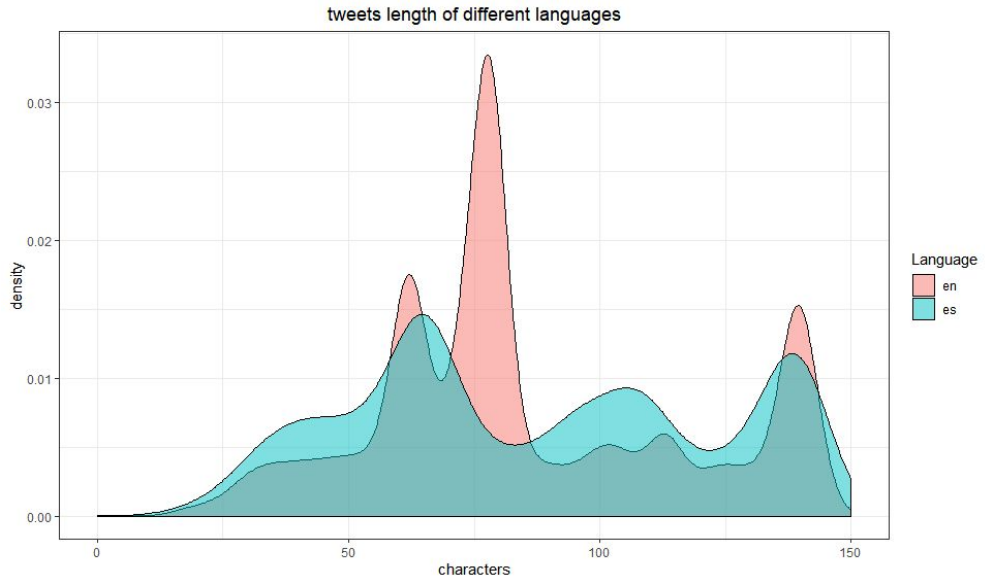
8.2 Two “Stars” with Strong Gravity in details



8.3 Additional insights from the data exploration

We learned the predominant two languages are “English” and “Spanish”. According to the techcrunch.com the limit of words is 280 characters and the most common length of a tweet is 33 characters. From the density plot below, we see the most tweets in English have 80 characters and most tweets in Spanish have 60 characters, meaning both languages are higher than the average word characters. Therefore, we assume users would have better interactions

and engagement with the game or other users during the game.



9. SOURCES

[1]https://www.statista.com/statistics/616373/nba-finals-cost-tv-commercial/
[2]https://marketing.twitter.com/na/en/insights/twitter-changes-the-live-tv-sports-viewing-experience
[3]https://www.kaggle.com/xvivancos/tweets-during-cavaliers-vs-warriors
[4]https://www.espn.com/nba/playbyplay?gameId=401034615