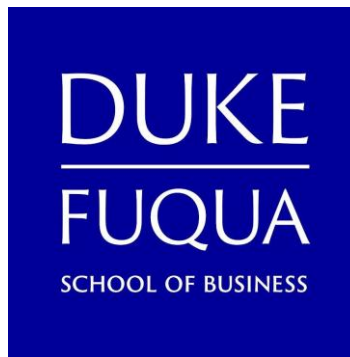


# **Duke University**

## **Sentiment Analysis Write-up**

### **MARKETING 549Q - Digital Marketing**



2019-10-16

## **I. Executive Summary**

The recent controversy regarding Houston Rockets manager Daryl Morey's tweet on supporting Hong Kong's protests on October 4th, 2019 has just escalated to wider conversations on Free Speech, National Sovereignty and Business Cultural Awareness. This one tweet burgeoned into NBA's worst nightmare. Even though Morey later apologized for his tweet, it set up a firestorm in China and later a counter-reaction in the US. NBA received a lot of backlash when they announced that "[The NBA] are not in a position to regulate what players, employees and team owners say or will not say", but they later issued a statement saying that "they will regulate the extent to which freedom of speech can be used." <sup>(1)</sup> The league has spent years and millions of dollars investing in China, an ever-growing market of 300 million basketball fans. <sup>(2)</sup> Morey's tweet led to a substantial damage on NBA's accessibility to Chinese market and Chinese-based business partners, as big companies such as Tencent and Li-Ning have cut ties with the Rockets and the NBA. There has been a mixed bag of sentiments from both sides of the aisle, and it is still a hotly debated topic on Twitter even 10 days later.

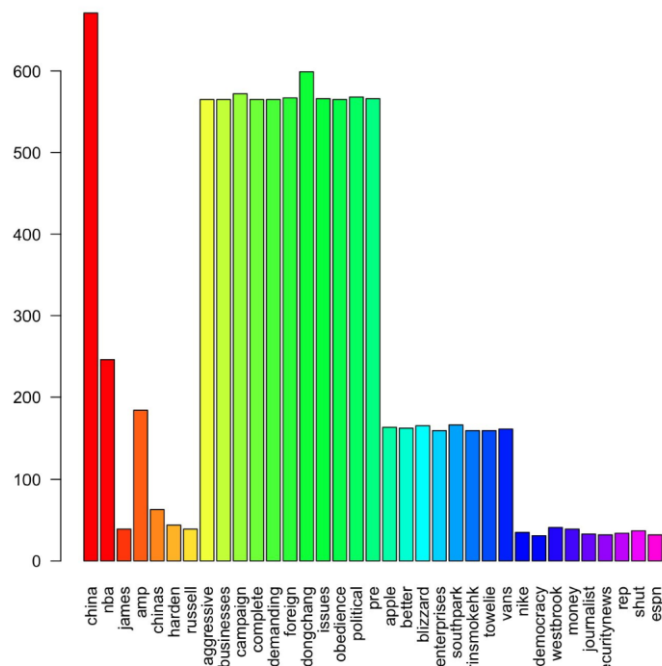
Based on 1,200 tweets from the hashtag #HoustonRockets pulled from Twitter.com at 11:23 PM on October 14th 2019, the public has strikingly polarized sentiments regarding the brand of Houston Rockets due to recent events. Most texts mined allude to a negative point of view. The word cloud presents topics in the realm of politics and business, both mainly concerning China. Social Network Analysis shows the NBA as a highly central node, connecting one cluster of other businesses relevant in the same conflict, and another cluster of political ideas relevant in the same conflict.

## II. Data Understanding

The brand that we chose is the NBA club that's in the center of the controversy - The Houston Rockets. Our analysis is based on the data extracted from real-time tweets. This topic was chosen because it is one of the most recent controversial issues which led to a variety of reactions. Since Twitter is blocked in China, our analysis is geared towards the sentiment from the rest of the world. We hypothesized that the prevailing reaction would be highly negative, and use text mining methods to test our hypothesis.

## III. Sentiment Analysis

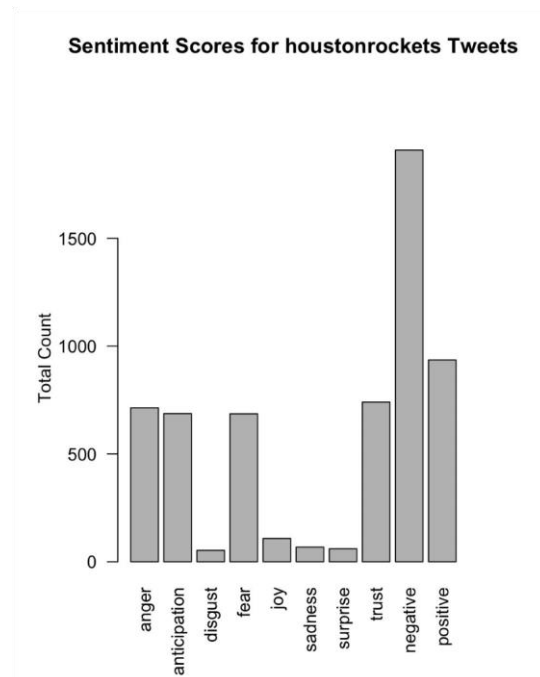
After cleaning and reformatting the storage of data, we created a bar plot for words that appear more than 30 times in the corpus. We iterated these bar plots to take out words that didn't provide much meaning for interpretation but weren't in the stopWords dictionary, such as "gets", "knows", "either", "like", and "can". The final clean set offered the following visualization.



We also put the document term matrix in a word cloud that signifies the frequencies of words and topics through sizes and text shades.



To further investigate the opinions regarding the Houston Rockets, we have compiled sentiment scores for the body of tweets based on the nrc\_sentiment dictionary, detailing eight emotions and two extremes: negative & positive. True to our hypothesis, a negative sentiment is more prevalent than positive. The public seems to view the issue regarding the Houston Rockets in a negative light, but there's also considerable prevalence for positive tweets.



However, dissecting the layers of the issue at hand, we knew that tweets on the Houston Rockets hashtag were directed at different entities: Morey's original tweet, the NBA's initial impassive attitude towards apologizing for Morey, the NBA's later apology, Chinese business retaliation, South Park's producers' ironic remorse statement, and other brands' actions amidst the conflict. It is hard to isolate the sentiment directed towards Morey's tweet from the part directed towards the NBA and towards other brands. Along with that, feelings of anger & fear and anticipation & trust both dominate, a divergent effect that makes the sentiment more complicated. Therefore, we decided to look into a widely retweeted tweet and inspect how components of the tweet measure against the sentiment directory.

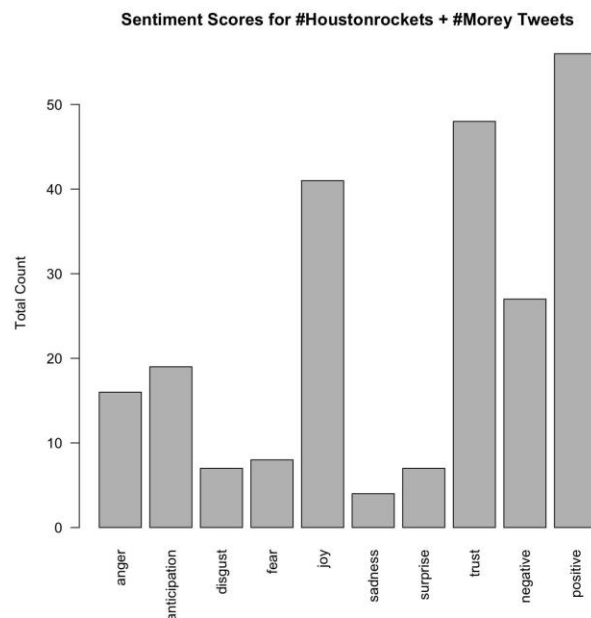


```
> get_nrc_sentiment('obedience')
anger anticipation disgust fear joy sadness surprise trust negative positive
1      0              0      0      0      0      0      0      1      0      1
> get_nrc_sentiment('foreign businesses')
anger anticipation disgust fear joy sadness surprise trust negative positive
1      0              0      0      0      0      0      0      0      1      0
```

While the term “obedience” scored 1 for positive and 1 for trust, “foreign business” was taken as 1 negative. As a result, the tweet counted towards both negative and positive, despite the tweet having an obvious sentiment of animosity to the human reader. In other words, the sentiment regarding the Houston Rockets is multifaceted and diverse: at both ends, opinions are extreme and passionate. Though the word cloud and sentiment score summary can help explain the situation, they are not optimal tools that can fully describe the specific layers of opinions regarding the NBA team at the moment.

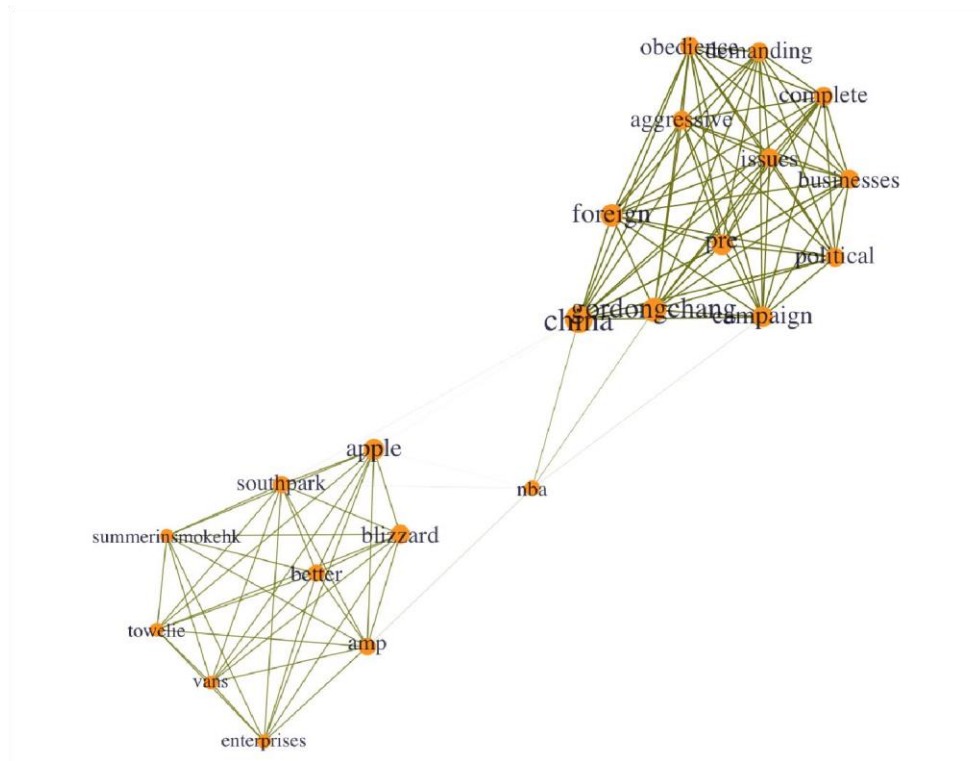
To mitigate this shortcoming, intuitively, we thought of combining two hashtags or more to narrow down the scope of the sentiment analysis. Unsurprisingly, tweets with both

#HoustonRockets and #Morey are relatively more positive than negative, with prominent feelings of joy and trust, which indicates that the public in question thinks positively of the Houston Rockets when it came to Morey. Repeating this task with other combination of hashtags will shed light on more layers of the issue.



#### IV. Network Analysis

Below is the visualization of the Houston Rockets hashtag network.



Similar to our observation in the sentiment analysis, there are two big clusters of topics: one encompassing political keywords, and one consisting of other businesses and brands. What we weren't able to identify from the word cloud is the NBA's position in the network. From the network, we saw that 'nba' as a node of great power, connecting a cluster of political topics and another of business topics. We dove further to analyzing its centrality qualities to better visualize the magnitude of its impact.

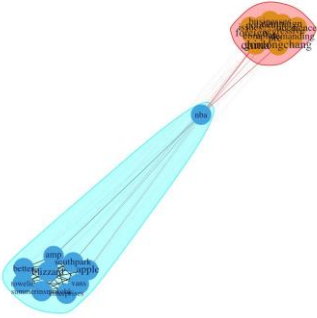
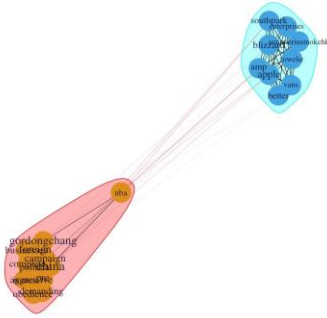
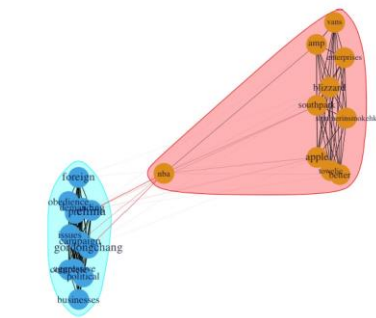
Node	Degree	Betweenness	Closeness	Eigenvector
nba	9	8.136447	0.01515152	0.4928238
china	16	46.961538	0.01923077	1
gordongchang	14	17.290293	0.01785714	0.9493824



apple	12	20.821062	0.01666667	0.4996167
-------	----	-----------	------------	-----------

Though ‘nba’ is an important node that connects two other big clusters, ‘china’ is the node with greatest degree, betweenness, closeness, and eigenvector. This means that ‘china’ is the most active node in the network (connected to the most nodes), has a lot of control (connects many nodes), is the closest to every other node, and is also connected to highly central nodes. This aligns with our exploration of the corpus: China is the most tied topic to the brand of Houston Rockets right now, and other subtopics in this hashtag are closely related to China. ‘nba’ isn’t actually the node with highest degree, nor betweenness, closeness, and eigenvector. This shows that the NBA has some distance between each of the clusters and doesn’t quite belong to either. However, it is still fairly connected with both groups, so the Houston Rockets, as a business, can’t really isolate themselves from any of the political and business influences with which they are involved.

Between deteriorating relationships with business partners abroad and disparate sentiment at home, the NBA and Houston Rockets have to devise a clear communication strategy to minimize damage, and also to establish a stern position on the matter. This strategy should be systematic and consistent from the director board of the association to employees of the team. One study they can do is the scope of each opinion’s impact on their overall business and brand value. They then can use this study for a cost-benefit analysis to make a decision that optimizes benefit, but also is true to their culture, mission, and values.

		
Edge Betweenness	Label Propagation	Greedy

Another interesting insight we found is that the three clustering methods categorize ‘nba’ differently. Edge Betweenness groups it with the business group, while the other two group it with the politics group. This further shows that the NBA is at the crossroads of compromising its business with political intents.

## V. Conclusion

Our text mining task returned four small groups of topics, which can be categorized into two big groups: politics and business. Most topics center around the controversy between China and Hong Kong, and how the Houston Rockets and other brands deal with it. A word cloud illustrated these topics in terms of frequencies. The sentiment analysis results were at first difficult to interpret, but when we investigated a combination of two hashtags, the sentiment became clearer. The network analysis delivered the same conclusion of segmentation: the overall sentiment is hard to grasp. Two extremes exist, with fervent feelings from both sides. We recommend that the Houston Rockets use our exploratory and unsupervised learning methods to conduct a cost-benefit analysis for future communication strategies.

### **Cited Sources**

1. <https://www.apnews.com/0a660e9e10664e31bf6ee359c22058cf>
2. <https://www.youtube.com/watch?v=OUBk5YQP128>