## Literature Review

The objective in this step to study available literature was to identify the problem scope and what social platforms to explore. To help with question construction we reviewed approximately four articles.

In the first article, we see that online cable has nearly doubled from 2015 to 2016. Growth in basic and broadcast cable actually declined. The second article, we see that the Foxtel service still retains its original market share but online media such as Netflix has increased by 30% from 2016 to 2017.

## Extension to assignment - Reddit Analysis

### Data Collection

Data is collected from Reddit using Python API, Tweepy. Total XX subreddits are collected in sequence manner. Based on the study, I used the keyword streaming and streaming videos to study the trends.

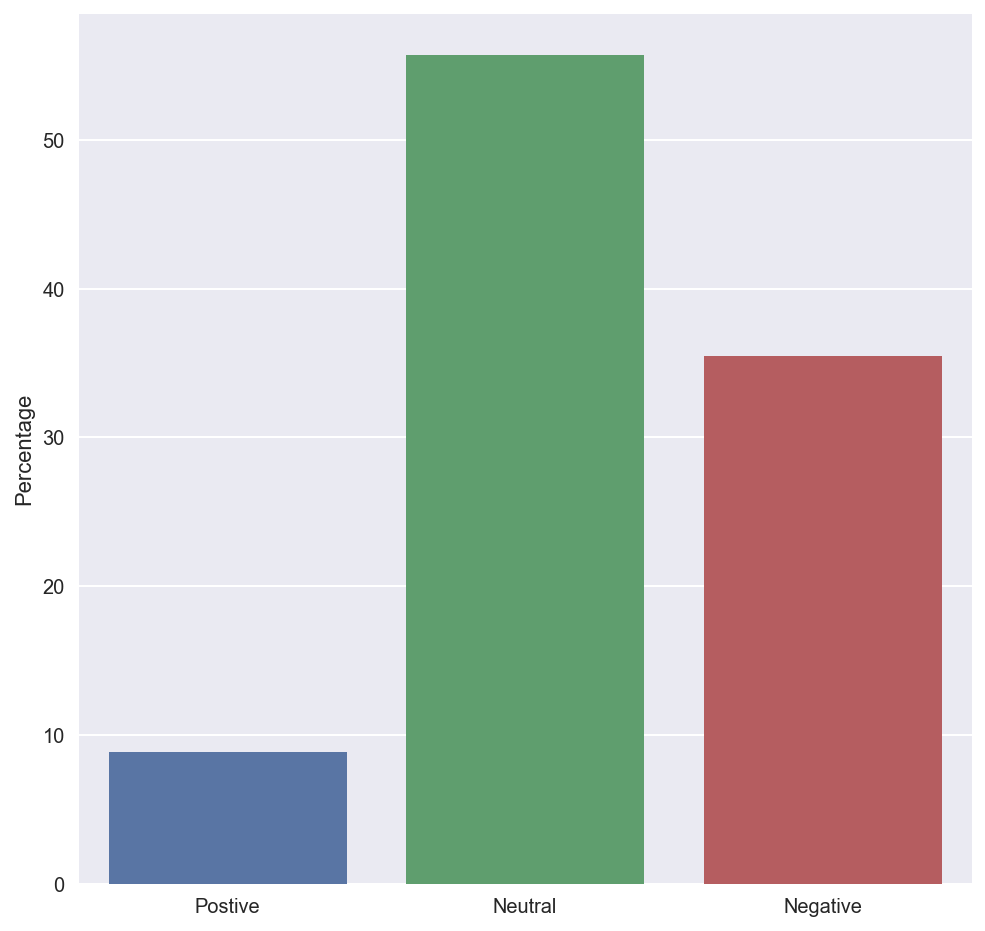
Code are implemented in Jupyter Notebook, Python 3. Pandas, Matplotlib, ntlk, networkx and seaborn are used for visualization purposes. Most of the code is using python in-built data structures such as lists and dictionaries.

# Analysis

After gathering and pre-processing the subreddit headlines, sentiment analysis is performed to obtain general overview.

## Sentiment Analysis

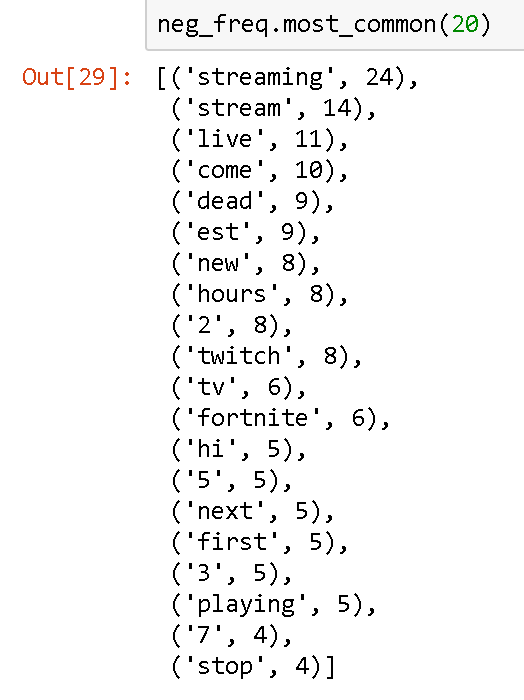
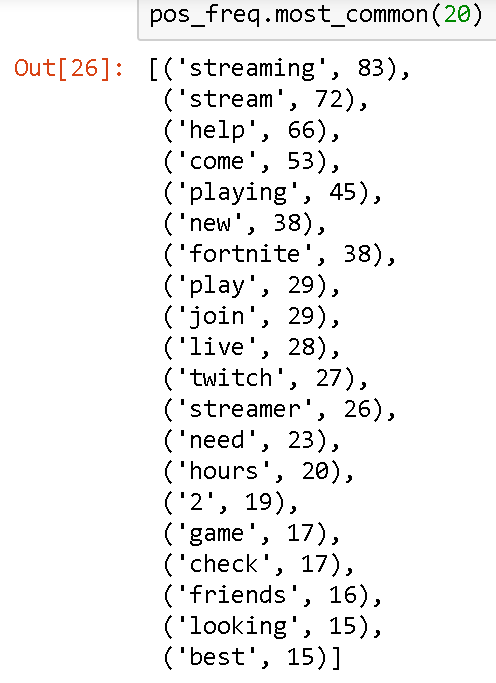
Below is the visualization of direct word match sentiments. Negative value in blue colour indicated negative tweets count, 0 is neutral and red is positive sentiment count. Distribution is not very symmetrical but this may be biased due to the query used.



*Figure 1 - Sentiment of subreddit posts*

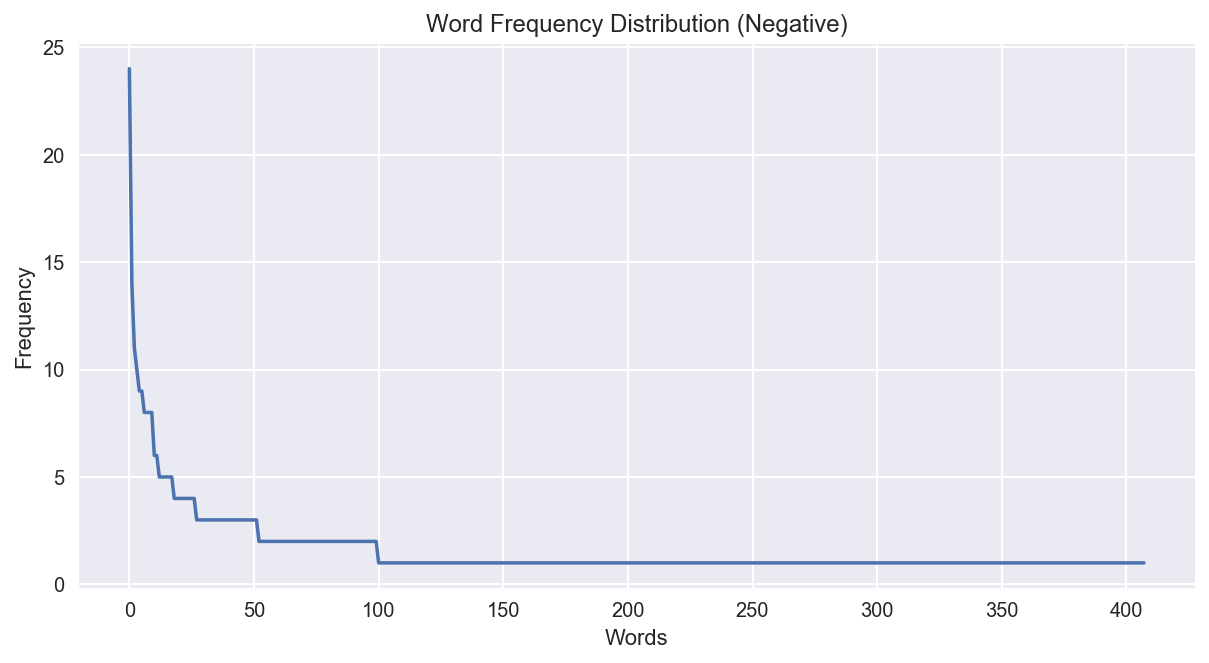
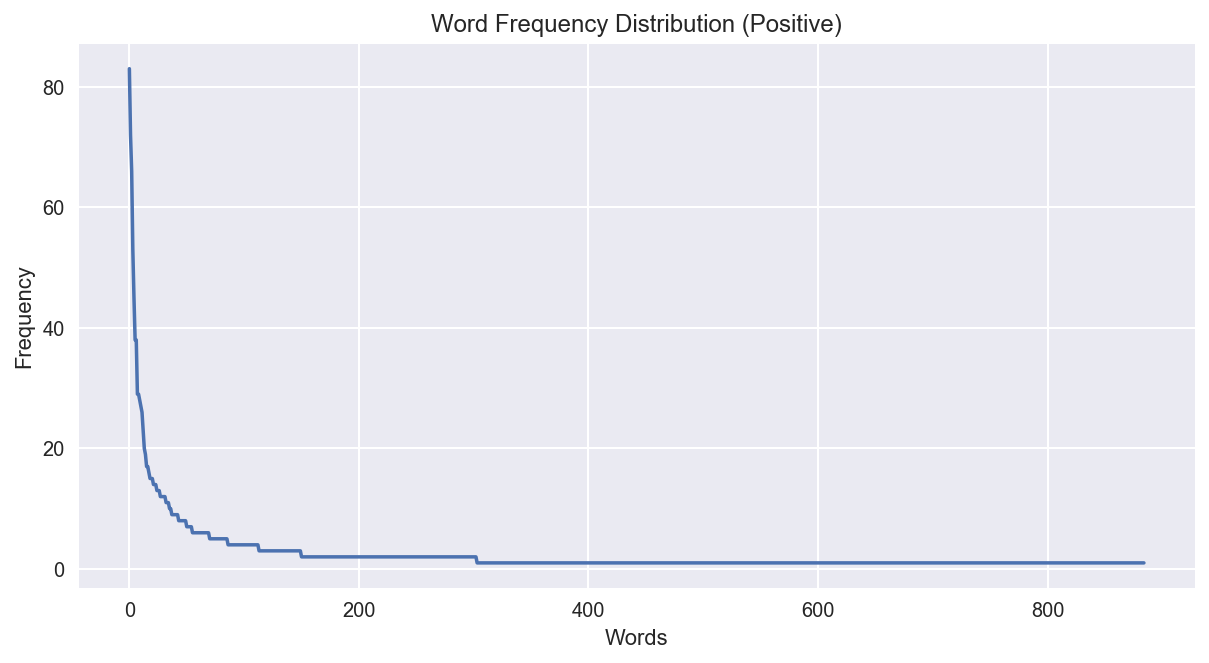
We investigated sentiment polarity using NLTK sentiment analyser module to get the compound scores.

Next, we’ll analyse the frequency of words used in reddit streaming headlines. Below we see the top 20 most frequent positive sentiment subreddits.



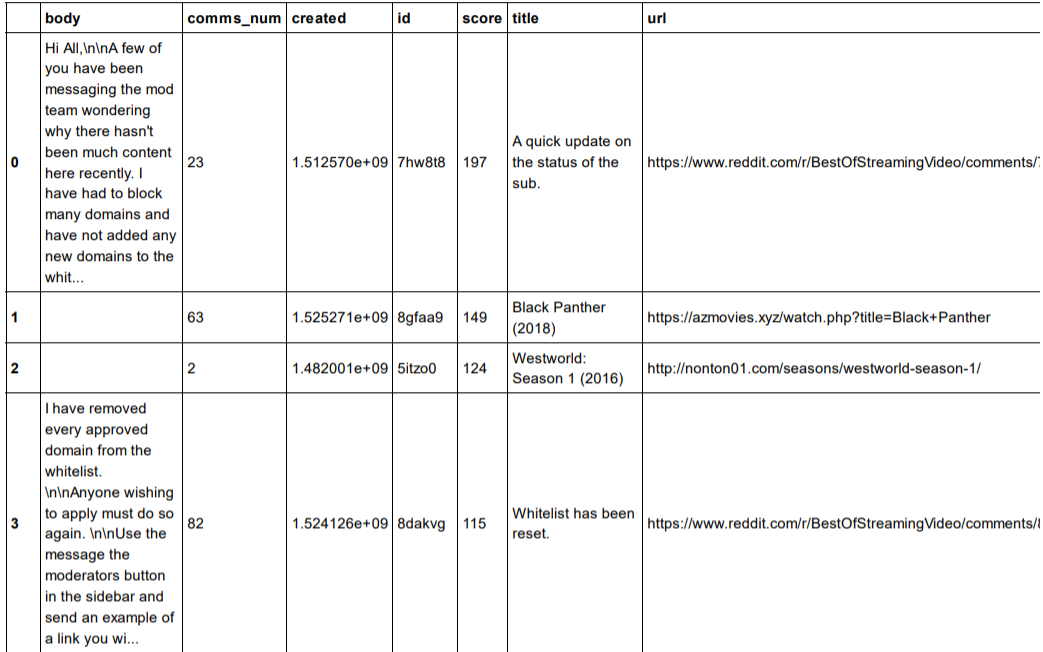
*Figure 2 - Highest Frequency Subreddit Headlines*

We can see that the frequency count for negative sentiment is considerably less than the positive sentiment.



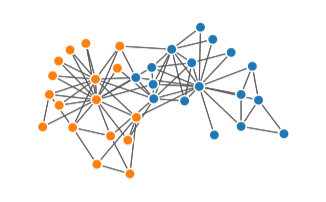
*Figure 3 - Frequency count of subreddits versus length of words in subreddits: Longer reddit posts are not viewed as frequently*

We then focused on streaming video and saw what was the sentiment attached to that. To analyse this we first created a topic dictionary of the top 500 posts related to streaming video where the subreddits were ranked by the score assigned by users who reviewed/visted the reddit posts.



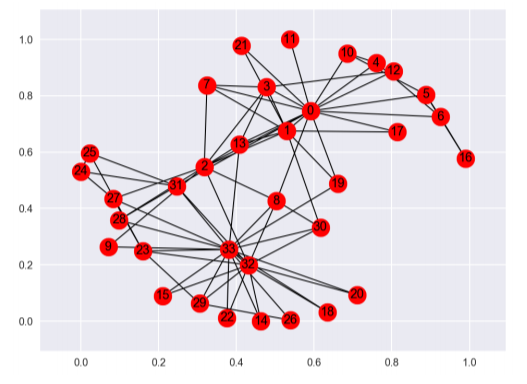
*Table 1 - Top Reddit user scored subreddits*

After the this we constructed a normal network diagrams.



*Figure 5 - Top Reddit user scored subredditsfor streaming video with modularity of 1.365 x 10-4*

We also constructed a spectral network diagram and found the network density to be approximately 0.008, which indicates a low interconnectivity between topics.



*Figure 6 - Top Reddit user scored subreddits spherical node diagram*

# Bibliography

Arlen, G. (2017, April 1). The Lines Blur Between Online and Broadcast. *TV TECHNOLOGY: INSIGHT MULTISCREEN VIEWS* , pp. 46-47.

Morgan, R. (2015). Netflix expanding market Foxtel steady in Australia. *Network Research Broadcast Media* , 50-62.