# Twitter Sentiment Analysis : Federal Election 2022

*An Analytical Prediction Based on Australian Labor and Liberal Parties' Twitter Sentiments and Networks Analysis*

Mohammad Wali Ullah (ulla0009), Darcy Sayer (saye0046), Adam Briers (brie0023)

**23 May 2022**

**Abstract**

***Social Media posts, including Twitter tweets, are increasingly used to express individual viewpoints, including political persuasions. The 2022 Australian federal election outcome will determine which of the two major political parties will run the country for the next term (3 Years). The study aimed to analyse the Australian general public's sentiment on tweets that are live pulled from Twitter to determine public opinion on the Labor and Liberal parties. A prediction of the expected election outcome based on this analysis has been drawn at the end of the report. The data was taken from Twitter based on the keywords and hashtags associated with the federal election and the two major political parties. Selected data were subjected to pre-processing and prioritised using exploratory data analysis methods. Finally, relevant data was manipulated using various tools like RStudio and Gephi to determine public sentiment toward the two major political parties. A more significant positive sentiment was indicated for the Labor party, from which we predict their procurement of the 2022 Federal Election. This result is consistent with the current polling conducted by Roy Morgan, Newspoll and YouGov.***

## Contents

Flinders
UNIVERSITY

# INTRODUCTION

The past 20 years can be characterised by the rapid growth of two entwined industries, large technological and social media companies. In 2004, the social media site Myspace hit one million active users, a never-before-seen feat. In addition, Facebook has 2.4 billion users, YouTube has 1.9 billion, and WhatsApp has 1.3 billion. Worldwide, 30% of the global population uses a social media platform, and in wealthy countries, up to 80% to 90% of adults engage with social media.

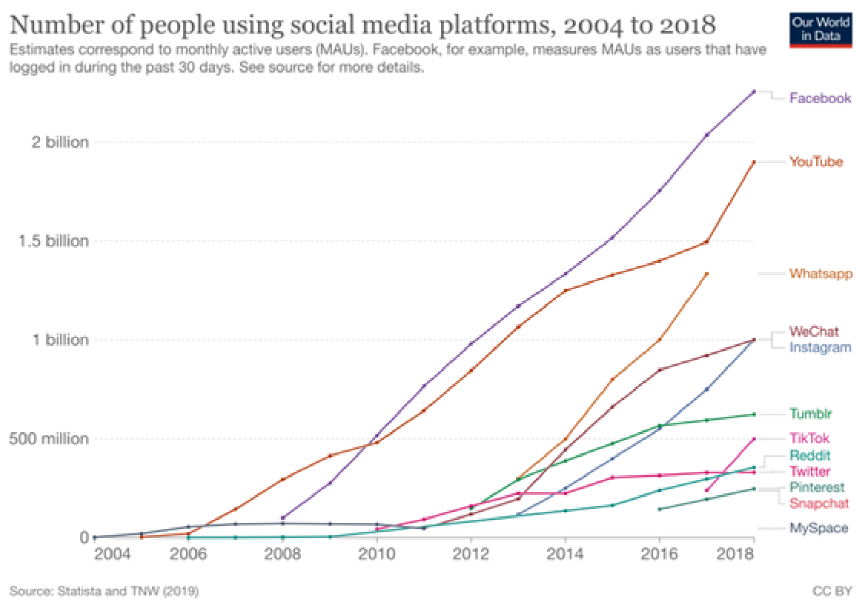


Figure 1: Social Media Users-Source: Statista and TNW (2019)

These companies profit from the astronomical quantity of data collected from their users, which has substantial economic value. In addition, data compiled by social media is sold to third parties, where it can be used for any number of reasons, be it targeted marketing campaigns or product development.

People use social media to voice their opinions on a variety of issues and topics, be it climate change, movie opinions or politics. A current topic of discussion is the Australian 2022 Federal Election, taking place on 21 May. Every Australian citizen 18 years and over is to vote on members of parliament, who are to represent their views in the House of Representatives. This election will decide whether the current Morrison government of the Liberal Party-led coalition will continue to run the office or if the opposition, the Labor party, led by Anthony Albanese, will be voted in. This result will have implications for the future of Australia and its Citizens. Hence it was the chosen topic for this research paper. Using data on the two major parties extracted from Twitter, a prediction is to be made on the election outcome.

Following this point, the Approach section describes the choices made to collect data from various sources and the tools used to manipulate and store the data. The Data Collection section specifies the data sources, the tools used to collect the data and the types of storage utilised. The Data Pre-Processing section details the raw data extracted and how it was cleaned to develop a corpus file and term-document matrix. The Exploratory Data Analysis section describes the early-stage manipulation and visualisation of the collected data to understand better and context of the online political landscape. The Twitter Network Analysis section compares the networks drawn from original datasets used to a much more recent snapshot of Twitter data. The Critical Analysis and Prediction section covers insights into all datasets, user networks and user timelines to draw further insights and predict the election outcome. Finally, the Conclusion section summarises the report's findings and recommends future research.

# APPROACH

In recent years, social media has provided an unprecedented channel of communication within the political landscape, including an abundance of platforms for users to converse and contest political opinions and allows political parties to leverage political campaigns, processes and activities. Social media websites such as Twitter provide a platform for a broad level of political discourse between users, and those involved in political processes may commonly use Twitter as a medium to mobilise and engage voters. However, the power of social media as a communication tool in the political sphere as to whether analysis of extracted content from websites such as Twitter may be a helpful tool to inform likely election predictions. Paired with the relevance of social media in the current climate of political discourse, this paper will focus on examining the ability of content shared on Twitter to be analysed and forecast a likely outcome for the upcoming Australian 2022 federal election.

With respect to the choice of data collection, Twitter is pivotal in the channel of political communication (Parmelee & Bichard, 2011). Indeed, in 2019 almost 80% of political members of parliament within Europe were found to use Twitter within their political campaigns to communicate with potential voters (Daniel & Obholzer, 2020). As a result, research has found that a level of public sentiment expressed by voters on mediums such as Twitter may successfully reflect accurate election predictions (Bansal & Srivastavaa, 2018; Tumasjan et al., 2010). Thus, the use of Twitter data content pertaining to the two major parties in the upcoming 2022 Australian federal election was selected as the data of choice within this paper.

Taken together, this research paper aimed to analyse the Australian general public's sentiment on tweets to determine public opinion on Labor and Liberal parties. The data extracted would then be used to determine the election's outcome.The latest political forecast polling has favoured the Australian Labor party as the preferential candidate in the upcoming Australian election from Feb-May 2022. It was hypothesised that the data would show a more significant positive trend to the Labor party compared to the Liberal party.

# DATA COLLECTION

Twitter data can be collected using various platforms such as **tweepy** in python, **twarc2** from CLI, the **RStudio** console, and the **Gephi** app. Our code works for all four of these IDEs. However, we have downloaded our Twitter data using **twarc2** and saved it in **.jsonl** format for this project. For the convenience of the analysis, the downloaded **.jsol** file has been converted to **.csv** format.

We first downloaded 5k+ tweets each for the Labor and Liberal parties on 12 April 2022, when the Australian Election Commission (AEC) declared the date of the Federal Election. We were interested in seeing how it changes over time as the election approached. Therefore, we downloaded another sets of 5k+ tweet for both parties on 30 April 22. To get a more accurate representation of the data, we combined the datasets and analysed 10k data for each party. At the end of the report, we indicate how the two big parties performed in the election campaign with the current leadership.

Mentioned below are the codes for data download using various platforms.

## Download twitter using twarc2

Download as *.jsol format

twarc2 search –limit 5000 "AustralianLabor" laborA.jsonl

convert *.jsonl to* .csv format

twarc2 csv laborA.jsonl laborA.csv

## Download from RStudio console

Creating token

token <- create_token(app = "....", consumer_key = "....", consumer_secret = "...", access_token = "...", access_secret = "...")

library(rtweet)

df <- search_tweets("@ScottMorrisonMP", n = 5, include_rts = FALSE, retryonratelimit = TRUE)

save_as_csv(df, "df.csv", prepend_ids = TRUE, na = "", fileEncoding ="UTF-8")

## Download using tweepy (V 3.10) in python

Import libraries

from tweepy.streaming import StreamListener

from tweepy import OAuthHandler

from tweepy import Stream

import twitter_credentials

Twitter streamer

class TwitterStreamer(): #Class for streaming and processing live tweets.

```
def __init__(self):
    pass

def stream_tweets(self, fetched_tweets_filename, hash_tag_list):
```

#This handles Twitter authentication and the connection to Twitter Streaming API

```
    listener = StdOutListener(fetched_tweets_filename)
    auth = OAuthHandler(twitter_credentials.CONSUMER_KEY, twitter_credentials.CONSUMER_SECRET)
    auth.set_access_token(twitter_credentials.ACCESS_TOKEN, twitter_credentials.ACCESS_TOKEN_SECRET)
    stream = Stream(auth, listener)
```

#This line filter Twitter Streams to capture data by the keywords: stream.filter(track=hash_tag_list)

Twitter Stream listener

class StdOutListener(StreamListener):

#This is a basic listener that just prints received tweets to stdout.

```
def __init__(self, fetched_tweets_filename):
    self.fetched_tweets_filename = fetched_tweets_filename

def on_data(self, data):
    try:
        print(data)
        with open(self.fetched_tweets_filename, 'a') as tf:
            tf.write(data)
        return True
    except BaseException as e:
        print("Error on_data %s" % str(e))
    return True

def on_error(self, status):
    print(status)
```

if **name** == '**main**':

#Authenticate using config.py and connect to Twitter Streaming API.

```
hash_tag_list = ["#AusVotes"]
fetched_tweets_filename = "AusVote.csv"

twitter_streamer = TwitterStreamer()
twitter_streamer.stream_tweets(fetched_tweets_filename,hash_tag_list)
```

## Download using Gephi app

Gephi is a user-friendly app to stream and analyse live data. It also has the tools necessary to analyse local datasets. There are options in Gephi that let users to produce various statistical analyses. In this report, we mainly downloaded all our data using twarc2, but for the network analysis, we have used Gephi to visualise the user networks from the live Twitter stream.

### Data Storage

After downloading the data, we saved all metadata in the MongoDB database.All project members have access to the database.

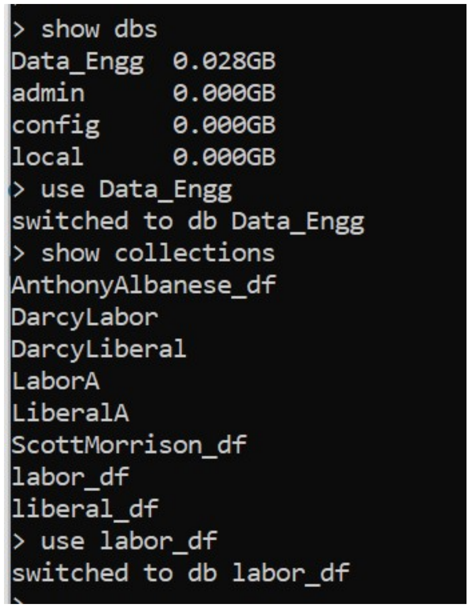


Figure 2: Accessing MongoDB from CLI

The **figure-2** shows that we can access the stored data from the command line interface. We also saved all the data into the local directory in **.csv** format. We then processed the data through various data pre-processing measures, creating a word Corpus and document matrix to use the data for analysis. Finally, our datasets are ready, and we will read the files from the local directory using the **read.csv()** function.

## DATA PRE-PROCESSING

### Build Corpus

We read the data from the local directory. We have used two sets of data, pasted both sets, and made a single set for each party. We then created two Corpuses. Codes are hidden in the pdf document. For more details, please refer to *.RMD* file is attached to this report.

### Clean text

All datasets have been cleaned to remove all unwanted words and characters at different stages. Codes are hidden. For more details, please refer to *.RMD* file is attached to this report.

# EXPLORATORY DATA ANALYSIS

In this section,we will create a word frequency chart based on the tweets to visualise the frequently used word. For the analysis, we have used **R** to conduct tailor-made based on what we wanted to extract and visualise from the data. Throughout the report, we utilised the analytical tools of **R**.We also used MS Excel to model the likes & favorite trends. The Gephi was used to analyse the social network of the target screen name and rendered the image in the report.

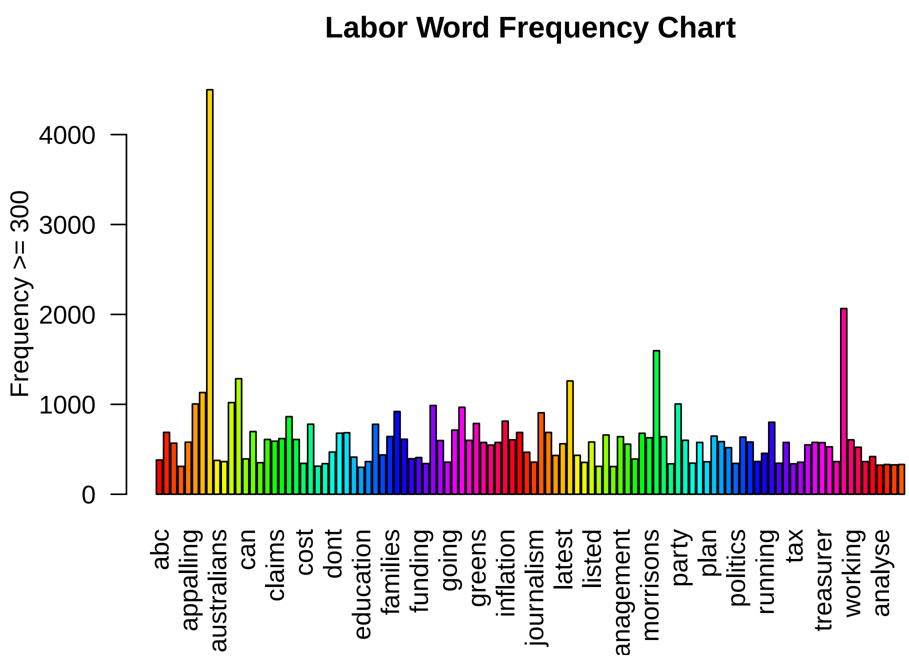


Figure 3: Labor Word Frequency Chart

Labor word frequency chart (**figure-3**) represents the words most frequently used in the 10000 tweets containing the **AustralianLabor** keyword. Keywords such as *Australians* and *working* are related to the party's identity and voting base. The chart also shows that some of the topics focused on are the **ABC**, **education**, **tax**, and **inflation**. Other keywords such as **appalling**, **don't** and **can** can be attributed to emotions and later used for sentiment analysis.

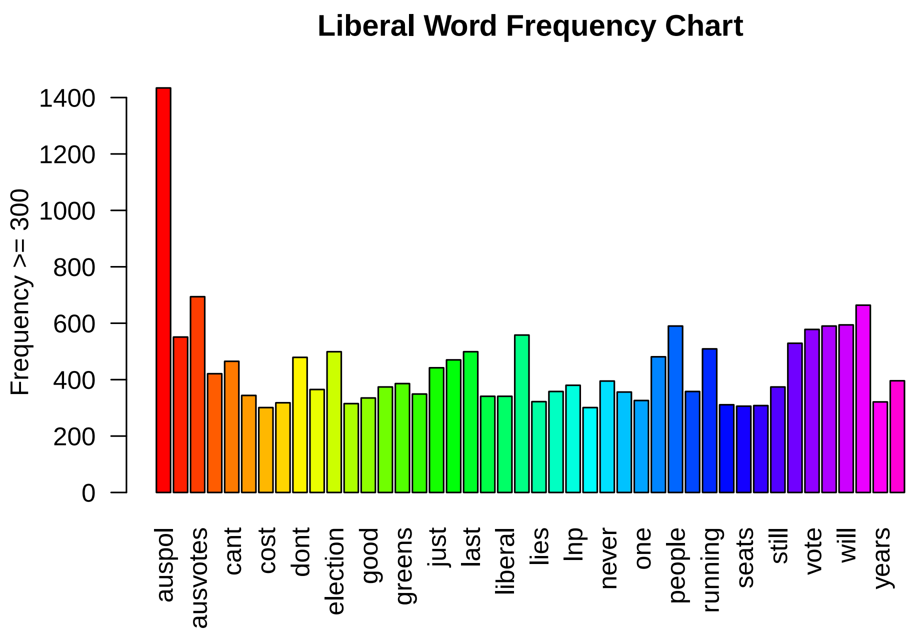


Figure 4: Liberal Word Frequency Chart

Liberal chart (**figure-4**) represents the words most frequently used in the 10000 tweets containing the **LiberalAus** keyword. Keywords focus primarily on the *election* itself, and topic discussion does not relate to significant issues. Keywords such as **can't**, **don't**, **good**, **lies**, **never**, **still**, and **will** all indicate more emotional tweets based on the liking/disliking of political figures rather than governmental issues. **Figures-2** visualise topic discussion about the Labor and Liberal parties, a necessary component when drawing comparisons between the two major parties.

When comparing the two figures, the topic discussion surrounding the Labor party focuses on **journalism**, **education**, **tax** and **inflation**.There is little mention of the election itself, especially when compared with the topic discussion surrounding the Liberal party. The majority of the words on the Liberal word frequency chart are emotionally charged or related to the election.

## Word Cloud

Word cloud is also used to visualise the concentration of the topic of discussion on social media. We will use the word cloud library to render the word cloud.

### Labor Word Cloud

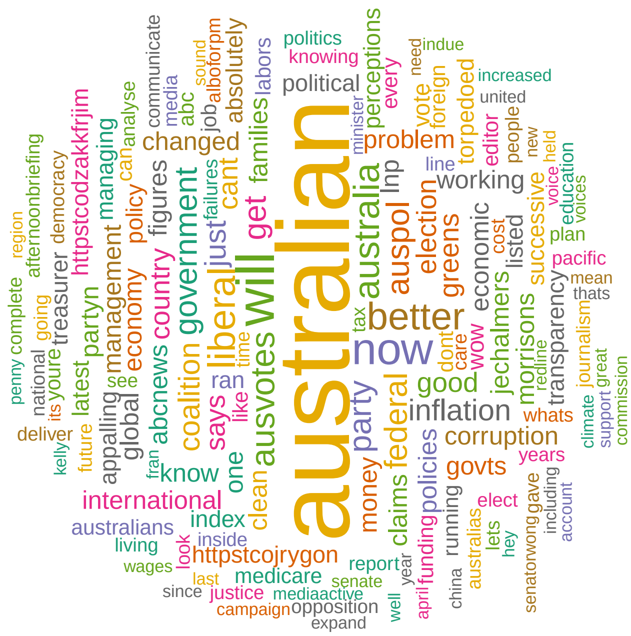


Figure 5: Labor WordCloud

**Figure-5** shows the concentration of topic discussion on Twitter surrounding the Labor Party with a lower word count requirement, meaning a more extensive range of words is displayed. As mentioned previously, topic discussion surrounding the Labor Party focuses on popular issues. This is further demonstrated here with more keywords such as Medicare, corruption, wages and commission.

### Liberal WordCloud

**Figure-6** shows the concentration of topic discussion on Twitter surrounding the Liberal party with a lower wordcount requirement, meaning a more extensive range of keywords is displayed. The increased scope shows that topic discussion surrounding the Liberal party does have some relation to issues with the inclusion of keywords such as corruption, climate and manufacturing.

Word clouds provide an alternative visualisation of the data for ease of understanding and allow for a broader scope of keywords due to lower word frequency requirements. Figures 3 and 4 build upon what has been previously mentioned and provide greater insight into the issues being discussed around both major parties.Note that we then pulled another 5000 tweets for both parties and performed comparisons of the most popular keywords.


Figure 6: Liberal WordCloud


Figure 7: Labor Comparison WordCloud

**Labor Comparison WordCloud 12.04.22 Vs 30.04.22**

**Figure-7** shows a word cloud with the most frequent keywords surrounding the Labor Party on 12 April vs the 30th, 18 days later. As the election draws closer, online conversation has primarily remained focused on various previously mentioned issues. However, there has been an increase in the frequency of words focused on the election itself. The more recent dataset contains keywords such as vote, votes, and running, which can be attributed to the increased discussion surrounding the election. These keywords are not observed in the older dataset. This comparison shows the change in topic focus surrounding the labor party as the election draws near.

**Liberal Comparison WordCloud 12.04.22 Vs 30.04.22**


Figure 8: Liberal Comparison WordCloud

**Figure-8** shows a word cloud with the most frequent keywords surrounding the Liberal Party on 12 April vs the 30th, 18 days later. As mentioned previously, the older dataset. Most frequently uttered words show the focus of the liberal party on women, treatment, Craig Kelly MP, who has been permanently banned from social media for the online distribution of "seriously misleading" information about Covid-19 vaccines.

Interestingly, after 18 days of the campaign, liberals drifted and focused more on funding, grants, etc. Using the "word-cloud2" library in R, we can create a clickable word cloud that allows the user to see which words were used more frequently and how many times.

**wordcloud2 (Labor-Top, Liberal-Bottom)**


## Sentiment Analysis

A sentiment analysis from our extracted twitter data will enable us to determine the feelings from the public regarding the Federal Election 22. By doing this technique it will determine if the data that has been extracted will be positive, negative or neutral. We must convert the tweets into character vectors using iconv() in R to perform the sentiment analysis.

### Labor Sentiment

```
##   anger anticipation disgust fear joy sadness surprise trust negative positive
## 1   208          158     156  232 106     196       86   252      460      412
## 2   245          192     204  280 160     240      103   316      590      544
```

```
##   anger anticipation disgust fear joy sadness surprise trust negative positive
## 1     1            0       1    1   0       1        0     0        1        0
```

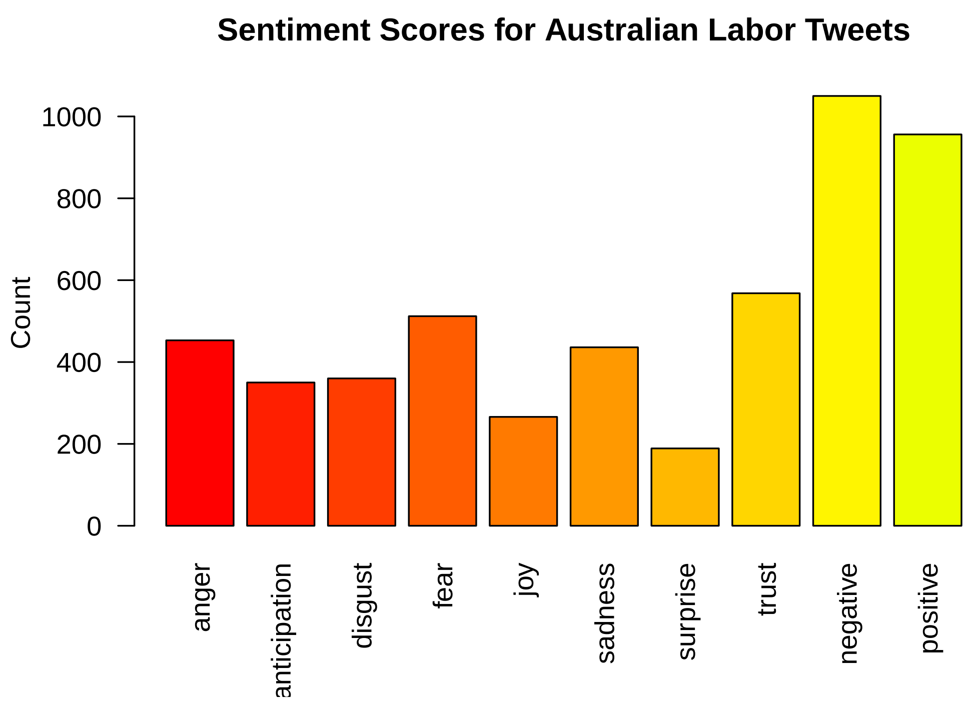


Figure 9: Labor Sentiment Score

Keywords in tweets are assigned an emotion, and the total count is stored in character vectors. For example, this column graph displays the total count of each emotion for discussion surrounding the Labor party.

**Liberal Sentiment**

```
##   anger anticipation disgust fear joy sadness surprise trust negative positive
## 1   298          226     230  334 175     269      119   337      682      579
## 2   260          200     227  275 156     239      109   309      618      527

##   anger anticipation disgust fear joy sadness surprise trust negative positive
## 1     1            0       1    1   0       1        0     0        1        0
```

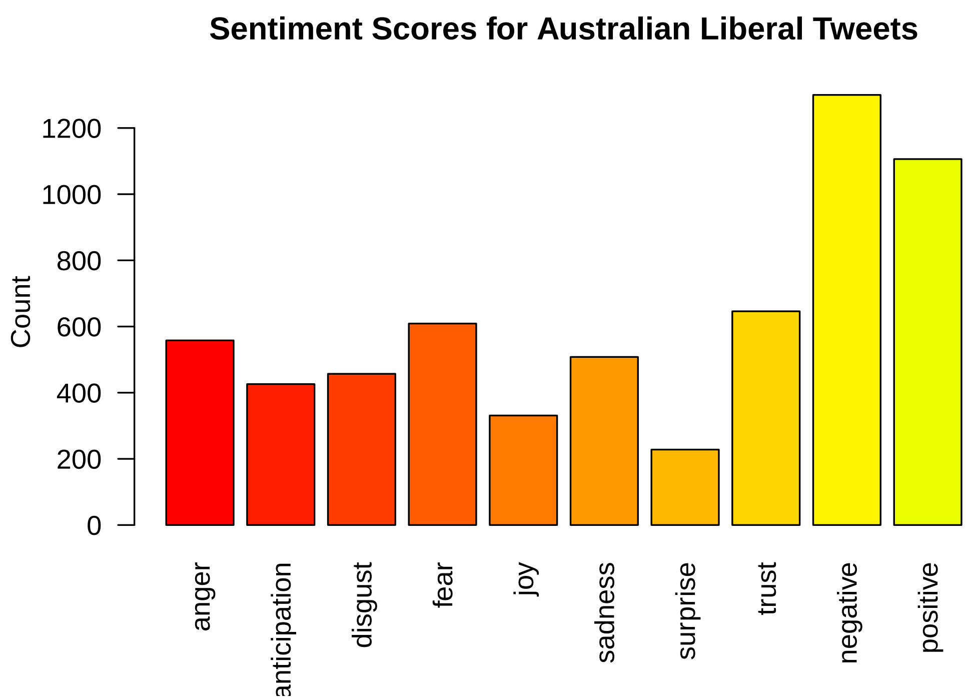


Figure 10: Liberal Sentiment Score

The sentiment analysis of the online conversations surrounding the labor party is crucial when making a prediction for the outcome of the election. Sentiment analysis can be used to gain an understanding of how the public views a given party at a given time. By conducting sentiment analysis on both the major parties we gain a snapshot view of public opinion on both Liberal and Labor parties. Comparing the sentiment scores of both parties, conclusions can be drawn on which party has the greater positive sentiment from the public. From this comparison a prediction of the election outcome can then be drawn.

## Network of Terms

To discover the communities of our data in the twitter network, this report analyses the community detection. The basic concept behind this test is to see the structure and methods of the clusters. We have chosen the most frequent words in political tweets to categorise the cluster. We then compiled, filtered, and visualised it for simple interpretation. To do so, we have streamed the timeline twitter data of both the labor and liberal party of Australia.

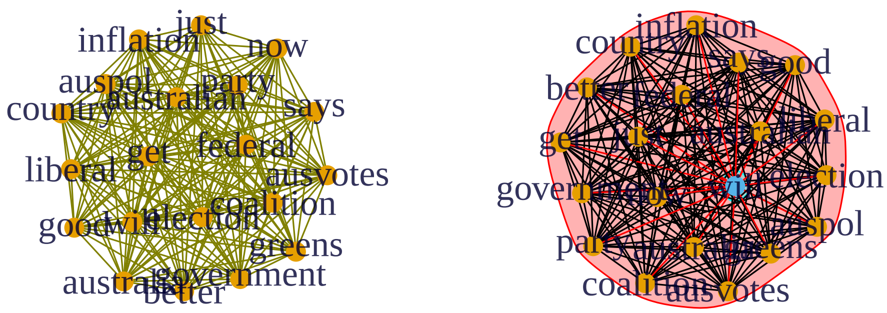


Figure 11: Labor Network Terms and Community Detection

**Labor Network Terms and Community Detection**

**Figure-11 -** The center focus of the labor community cluster is ‘will’. To visualise the community clusters, we have chosen word frequency >700. The most significant clusters are will, ausvotes, govern, get, better, inflation, justice, country, party etc. However, mentions of auspol, coalition, Australia Greens, Australian, and election also form a significant portion of entire clusters. Communities in our data appeared in multiple clusters.

It is observed that there is two distinct clusters group of clusters. We can see a small set of nodes connected with a more considerable number of nodes. All scattered clusters are again related to the hub cluster. It may be because our screen name was the name of the political parties. The political and social media account usually are run and managed by a dedicated person or a group f person. Therefore, the originator of the tweets is a small cluster compared to the followers’ communities and those who retweet & mention them.

**Liberal Network of Terms and Community detection**

**Figure-12 -** The centre focus of the liberal community cluster is women. To visualise the community clusters, we have chosen word frequency >450. The most significant clusters are women, know, votes, will, well, know etc. However, mentions of auspol, Australia, liberals, and elections also form a significant portion of entire clusters.

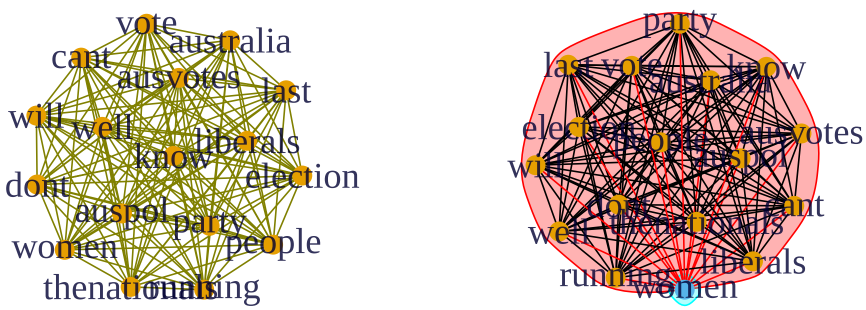


Figure 12: Liberal Network of Terms and Community detection

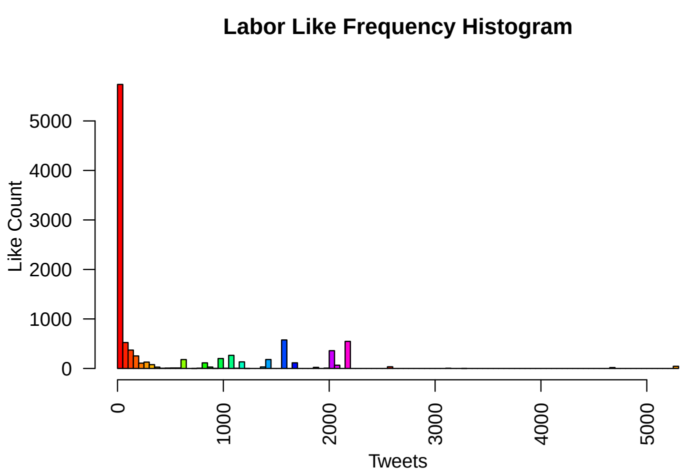


Figure 13: Labor Like Frequency

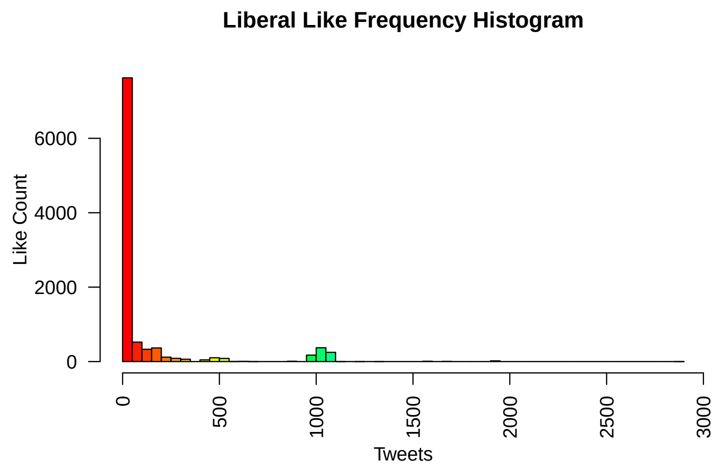


Figure 14: Liberal Like Frequency

**Figure-13** shows the likes count of labor tweets, and **Figure-14** indicates the liberal histogram of likes Vs tweets count. Australian labor is more engaging on the Twitter network than liberal Australia as a political party. When we analyse the state-wide tweets, according to **Figure-15**, it is also found that in almost every state & territory, labor tweets attract more likes compared to liberal. In particular, labor is more advanced in TAS, SA, QLD, and VIC.

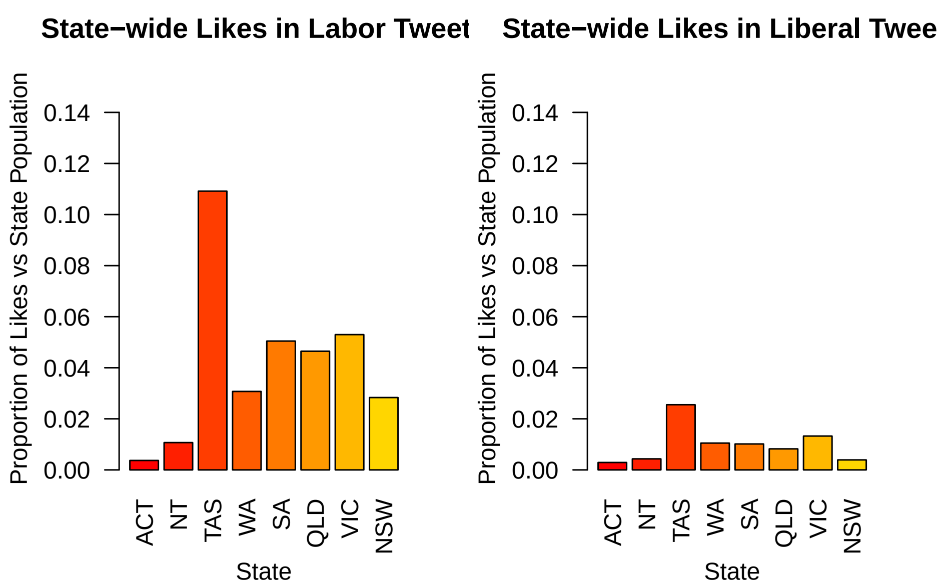


Figure 15: Like Vs State

## Favorite Count & Re-tweets Count

### Labor

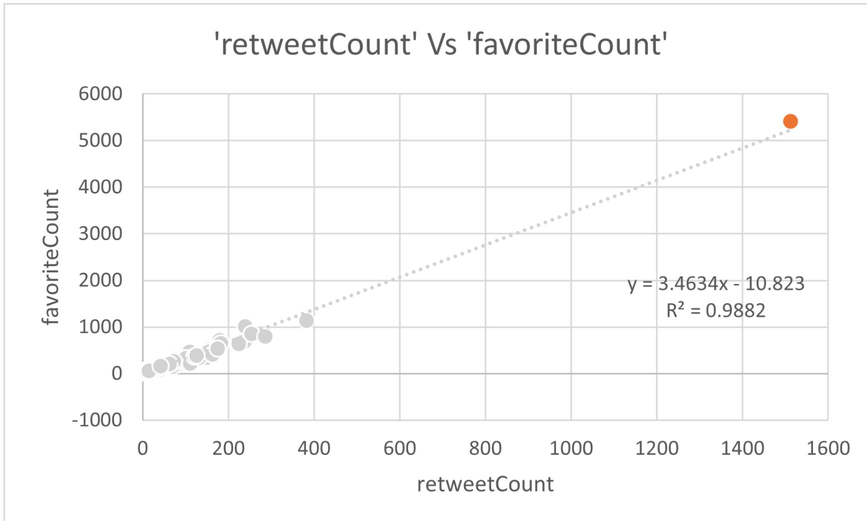


Figure 16: Labor FavoriteCount Vs RetweetsCount

Re-tweet count of labor against favorite count is highly correlated with 1 outliner. The R-squared, coefficient of regression determination, and value of labor data are significantly higher ($R^2 = 0.9882$). This linear regression model indicates that the labor party's tweets get a goodness-of-fit measure to correlate the retweet count & favorite count. The equation of the relation is y = 3.4634x - 10.823.

### Liberal

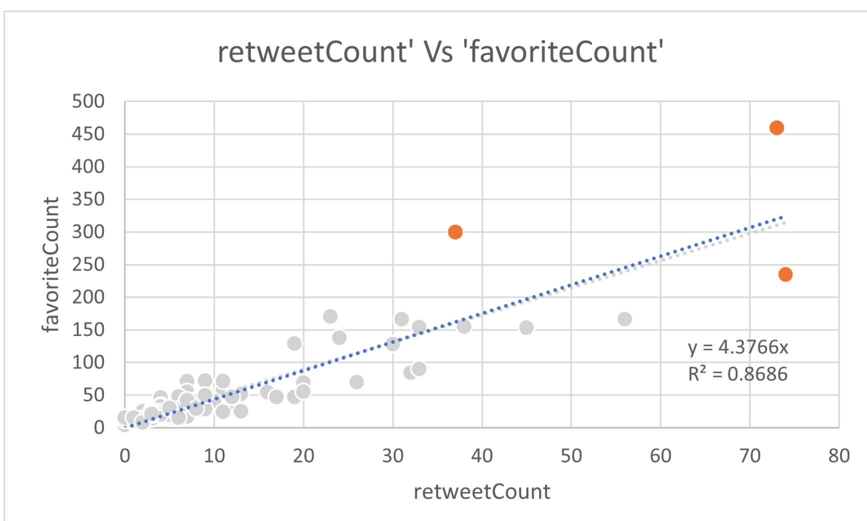


Figure 17: Liberal FavoriteCount Vs RetweetsCount

Retweet count of liberal against favorite count is highly correlated with three outliners. R-squared value($R^2 = 0.8686$)of liberal data is less significant than labor. The equation of the relation is higher y = 4.3766x. R-squared value is an indication but not necessarily the full-scale measure of goodness.

## K-Means Clustering

### Labor K-Means Clustering

```
## K-means clustering with 5 clusters of sizes 21, 18, 13, 22, 26
##
## Cluster means:
##   FollowersCount FriendsCount
## 1      0.8799310   1.31991024
## 2      1.3109333   0.91051552
## 3      0.5616613   0.67487764
## 4      0.2566448  -0.14317191
## 5     -0.1854233  -0.02351741
##
## Clustering vector:
##   [1] 5 5 5 4 5 4 5 5 4 5 5 5 4 4 4 3 4 4 5 4 4 5 4 5 5 5 5 4 5 5 5 4 4 5 4 4 5
##  [38] 5 3 5 5 4 4 4 5 4 5 4 5 4 1 1 2 1 1 3 1 1 1 2 3 2 3 1 2 2 1 1 2 3 1 2 2 3
##  [75] 3 2 2 1 1 3 3 1 3 1 3 2 1 2 3 1 1 1 2 2 2 1 2 1 2 2
##
## Within cluster sum of squares by cluster:
## [1] 1.456013 1.487897 1.030332 1.421574 1.813099
##  (between_SS / total_SS =  89.6 %)
##
## Available components:
##
## [1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
## [6] "betweenss"    "size"         "iter"         "ifault"
```

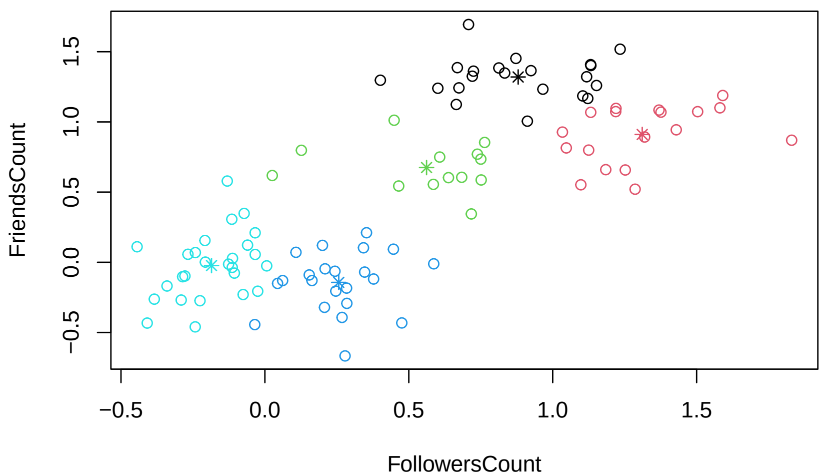


Figure 18: Labor K-Means Clustering

When we analyse the data, we see different patterns and clusters in data. However, using the k-means algorithm, we can transform data and visualize them with appropriate colors. Clustering centers give the clear pictures of respective clusters.Due to the size of data we have chosen to take random five centers.

We plot the labor clusters of friends count, and followers' count. For the labor K-means, the algorithm randomly chose 21, 18, 13,22 & 26 as pivots and calculated the mean values of both followers & friends.We can see the betweenness centrality of labor nodes is very significant (89.6%). It means labor betweenness centrality might have better control over the social network.

### Liberal K-Means Clustering

```
## K-means clustering with 5 clusters of sizes 20, 17, 16, 23, 24
##
## Cluster means:
##   FollowersCount FriendsCount
## 1      0.7611307  1.257834344
## 2      0.9307485  0.520930012
## 3      1.2258158  1.044682545
## 4     -0.2636738 -0.009038102
## 5      0.2237535  0.015341015
##
## Clustering vector:
##   [1] 5 4 4 4 5 5 5 4 4 5 5 4 4 4 4 4 5 5 4 4 5 2 5 5 5 5 5 4 5 4 5 4 5 5 4 4 5
##  [38] 5 5 2 4 5 4 4 5 4 4 2 4 5 1 1 2 3 3 1 2 2 2 2 1 1 1 2 3 1 3 1 1 3 1 1 1 2
##  [75] 3 3 1 2 3 1 3 1 2 2 3 2 3 1 3 3 1 1 1 3 2 3 2 3 1 2
##
## Within cluster sum of squares by cluster:
## [1] 1.7754915 1.5270036 0.9214638 2.0810656 2.2054797
##  (between_SS / total_SS =  86.8 %)
##
## Available components:
##
## [1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
## [6] "betweenss"    "size"         "iter"         "ifault"
```

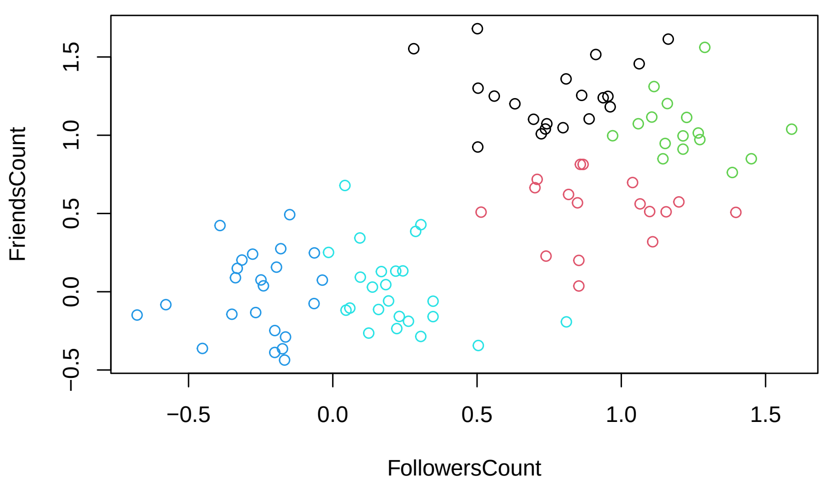


Figure 19: Liberal K-Means Clustering

Similarly, five random pivotal numbers of 20, 17, 16, 23 & 24 have been applied. Finally, the K-means algorithm calculated the mean values of liberal followers & friends. According to The Australian Financial Review report on 09 May 22, labor, a two-party preferred, ahead by 52%, whereas the coalition led by liberal 40%, with 8% undecided. Election experts opine that the RBA rate rise hurts the coalition.

The steady rise in labor alliance is observed through the labor K-Means clustering analysis of retweets and favorite counts. We can see our regression model in line with the newspaper report. Our analysis initially showed liberal a down-trending sentiment which gradually recovered within the first 20 days of the campaign. However, due to the recent RBA decision, it plunged liberal again. Through the news and print media, it is reflected that Morrison led the coalition to lose the personal support of the acting prime minister. Morrison's net negative is 19 points, and labor leader Anthony Albanese's net negative of 6 points. Every tool around the federal election 2022 indicates the noticeable popularity fluctuations for all major political parties.

Table 1: Tweet Count per Candidate

| candidate | tweet count | proportion |
|---|---|---|
| AlboMP | 923 | 65.46 |
| AustralianLabor | 135 | 9.57 |
| LiberalAus | 85 | 6.03 |
| ScottMorrisonMP | 267 | 18.94 |

### Combined sentiment analysis based on data extracted on 30 April 2022.

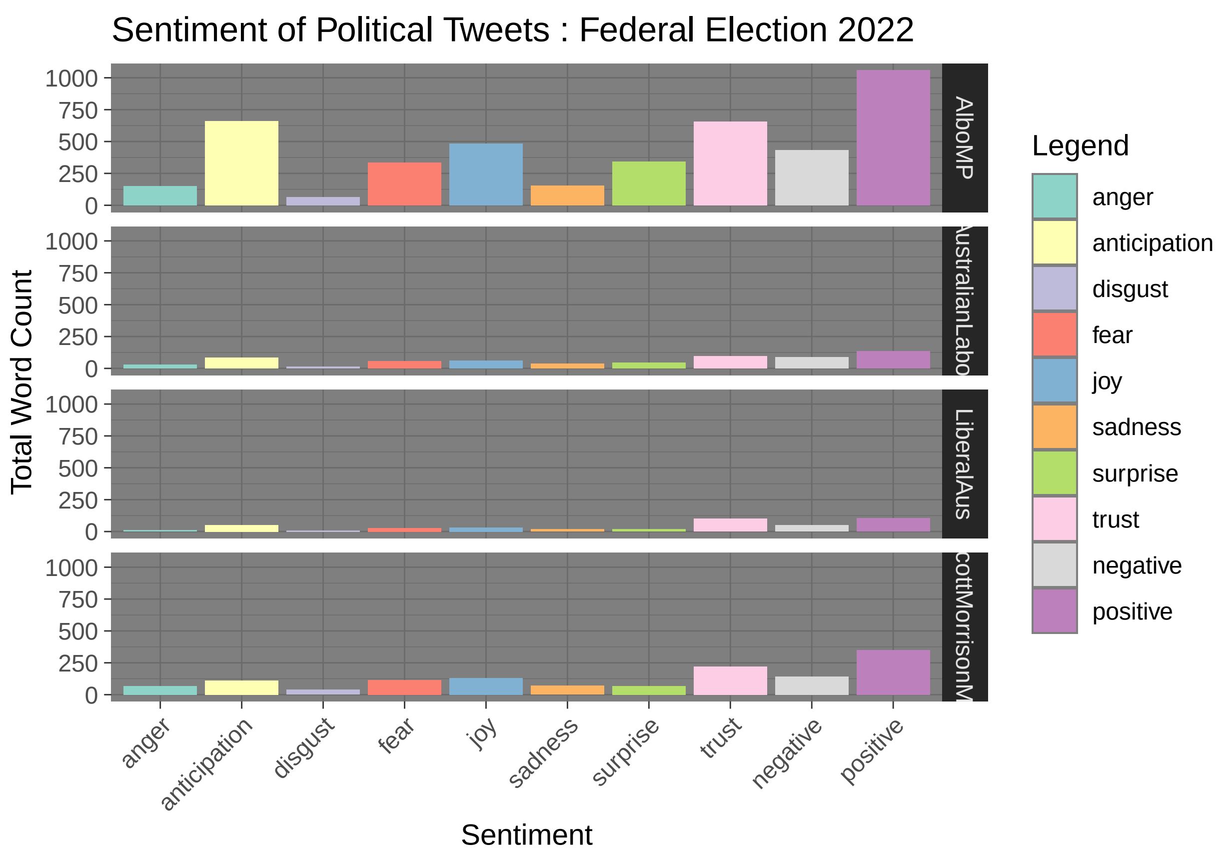


Figure 20: Combined Sentiment Analysis

Labor's online conversations remained focused on various previously mentioned issues such as Medicare, corruption, wages and commission, however, a few new focuses appeared such as vote, votes and running. Whereas in the later set of tweets Liberal's online presence seemed to have changed and headed in line with the election's topics. After Liberal's first set of data of keywords it was showing that their sentiment analysis was very negative, a big part of that would have been due to Craig Kelly MP being banned on social media platforms for sharing misleading information about Covid-19 vaccines. Once the second set of data had been analysed their sentiment analysis showed that Liberal was still negative, however their positivity increased quite significantly, with new keywords such as funding and grants.

Table 2: Most Frequently Tweeted Words in FederalElection2022

| words | freq |
|---|---|
| will | 225 |
| labor | 208 |
| government | 165 |
| australians | 147 |
| australia | 131 |
| plan | 125 |
| today | 114 |
| better | 110 |
| morrison | 107 |
| care | 103 |
| future | 84 |
| live | 73 |
| scott | 70 |
| australian | 67 |
| great | 67 |

## Word Cloud of Election 2022


Figure 21: Combined WordCloud

The word cloud of the election 2022 (**figure-21**) was generated from data pulled from the **"AustraliaFederalElection2022"** keyword. It was done to gain a clearer idea of the central points of discussion for the election without focusing on a particular political party. The word cloud exhibits the main points of the discussion centred on the COVID-19 Pandemic, Medicare, work security, and the price of living. It should be noted that the labor party is significantly more active on Twitter, and thus data will be skewed towards them.

When comparing Labor and Liberal word clouds, several observations can be made. First, there is a significant overlap between the labor party word cloud and the election word cloud. Both share topics such as Medicare, work and jobs. However, it should be noted that the election word cloud does not mention corruption or the environment, unlike the labor party word cloud.

# TWITTER NETWORK ANALYSIS

## Scott Morrison MP

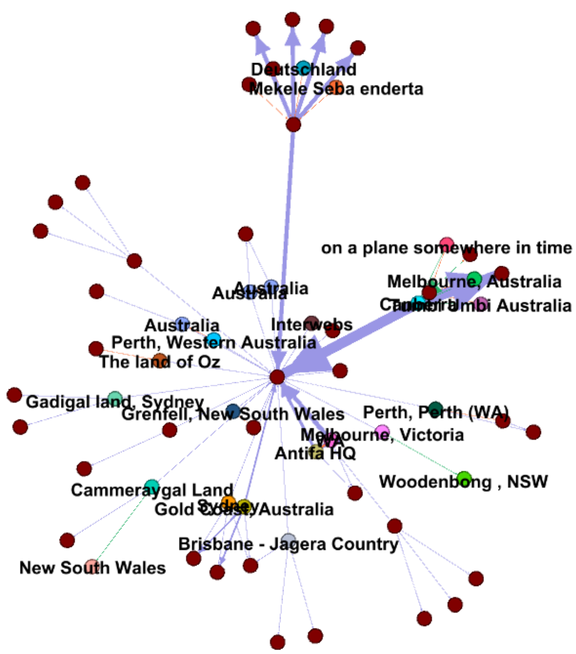


Figure 22: Scott Morrison User Network

Conversation on social media networks reflect the contours as people reply and mentions one another. Political tweets get polarised as they interact with recognisable groups. The **figure-22** shows the twitter network of Scott Morrison, Liberal Leader and acting PM of Australia. We have a live stream and captured 62 nodes with 100 edges. 32.2% of the tweets have been mentioned, and only 5.08% have been retweeted. Tweets occupy 18.64% of the network, of which 23.73% have a hashtag.

The centrality of the tweets at Sydney. It indicates that most of the tweets originated from Sydney and were mentioned from all over Australia. The highest mentioned has been observed from Melbourne.

### Anthony Albanese MP

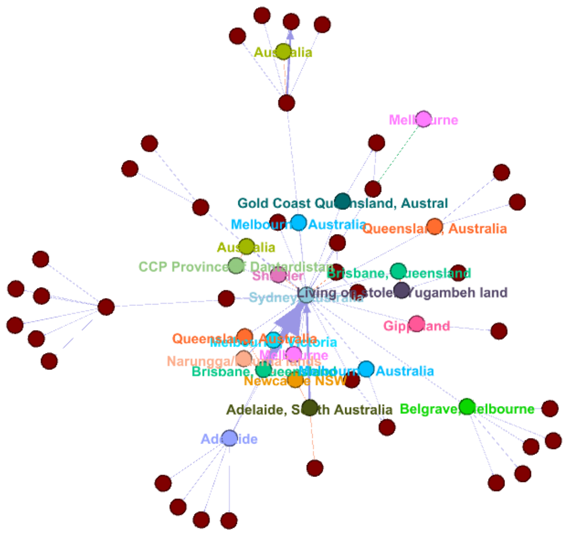

Figure 23: Anthony Albanese User Network

The **figure-23** shows the twitter network of Anthony Albanese, Labor Leader and acting Opposition Leader of Australia. We have a live stream and captured 60 nodes with 90 edges. 28.57% of the tweets have been mentioned, and only 12.86% have been retweeted. Tweets occupy 27.14% of the network, of which 11.43% have a hashtag. The centrality of the degrees at Sydney. It indicates that most of the tweets originated from Sydney and were mentioned from all over Australia. The highest mentioned has been observed from Melbourne.Comparing both leaders' Twitter networks, it is learnt that Albanese's tweets get less reaction from the followers than Morrison's. It is also found that the Opposition leader tweets more than PM.

### Federal Election 2022 Network

The **figure-24** shows the full networks of Federal Election 2022. From the live tweets, we have captured 50 nodes in 59 edges. It is found that Sydney and Queensland, among all big cities in Australia, are more engaging, contributing 20% of the total tweets network. Currently, 10.81% of tweets get retweeted instantly, whereas almost 85% get mentioned. Besides, a non-related cluster on the topic has been observed from outside Australia in Delhi, India. Due to irrelevant to our analysis, we ignored that small cluster. The figure also shows two non-linked clusters in Bondi, NSW, and another between Adelaide & Darwin. Interestingly, a more significant engagement has been observed throughout the Australian Indigenous communities' locations across Brisbane.

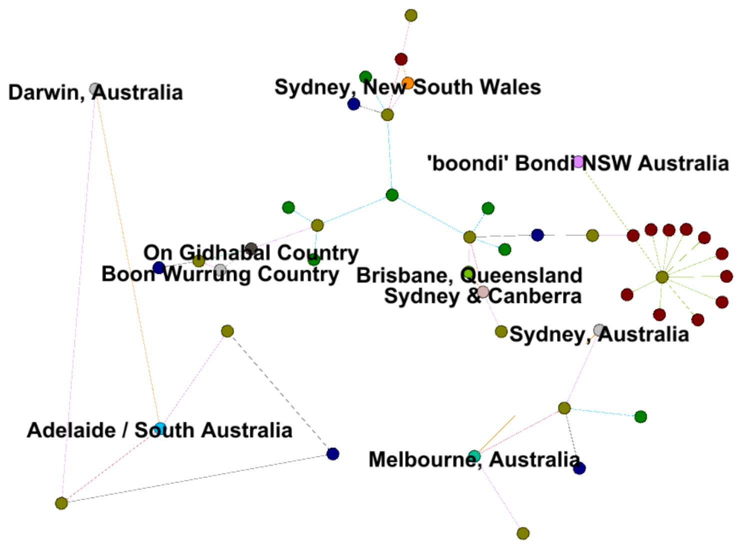


Figure 24: Federal Election 2022 Full Network

# CRITICAL ANALYSIS & PREDICTION

Analysis of the initial word frequency charts found that the topics of focus for both major parties were quite different. The Labor party focused on problems that the public feels are important such as the ABC, education and tax. On the other hand, the Liberal party did not focus on apparent issues. The most frequently used words by Liberal party associated tweets indicated more emotional statements focusing on public opinion of liberal political figures. Note that the woman and women treatment-related keywords are prevalent here; this is likely due to the recent reports of poor treatment of women in the Liberal party. One drawback of these visualisations was the large quantity of data being processed; the minimum word frequency needed to be 300 so the data could be readable.

Word clouds were generated with a lower word frequency requirement to broaden the scope and gain greater insight into what is being said. Findings from the Labor word cloud reinforced what was found in the word-frequency chart with more keywords surrounding topical issues such as Medicare, corruption, Commission and wages. The Liberal word cloud also increased discussion surrounding topical issues such as climate, corruption and manufacturing. It is also worth noting that one of the frequently tweeted keywords was craigkellymp. This was due to Liberal MP Craig Kelly being banned from all social media in Australia because of seriously misleading tweets regarding covid-19 vaccines.

The project team then created another two-word clouds comparing word frequency changes over time to draw more insights from the word frequency data. This allowed us to observe changes in the online political landscape as the election approached. The Labor comparison word cloud did not change for the most part. The online conversation focuses on the issues previously mentioned with an increase in keywords related to the election, such as vote, votes and running. The Liberal comparison word cloud showed a drift in focus from treatment of women in the liberal party and various opinions on liberal figures towards grants and funding.

One of the significant drawbacks of these comparison word clouds is that it is impossible to see how keywords have changed over time. The viewer has to read through and attempt to notice if words are present in one dataset and not in the other. Another drawback is that while we can observe that many of the keywords have emotions associated with them, sentiment towards a party is not clear.

To give a clearer picture of the sentiment towards both parties, sentiment analysis was performed on the cleaned and compiled datasets and then visualised in bar chart form. The Labor sentiment bar chart showed that negative emotions such as anger, fear, sadness and disgust had higher counts than positive emotions such as anticipation, joy, surprise and trust.

It was also observed that the negative sentiment had a slightly higher count than the positive. While this would initially lead the viewer to believe that the Labor party would not be favored to win the election compared with the Liberal party sentiment, the opposite becomes apparent. The Liberal party sentiment chart showed a significantly higher score for anger, disgust, fear and negative sentiments than the Labor party. The total emotion count was also much higher than the Labor party; this aligns with earlier insights that the Liberal tweets were much more emotional and opinion related than the Labor Tweets.

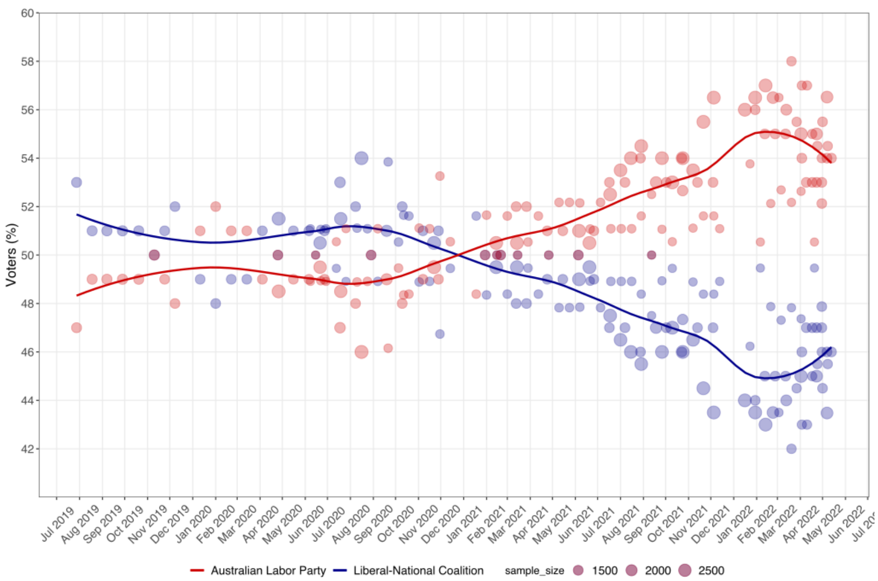


Figure 25: Opinion polling for the 2022 Australian federal election-Extracted from Wikipedia

Research into what could be causing this increased negative sentiment towards the Liberal party revealed several potential causes. Liberal MP Craig Kelly is likely a cause of increased negative sentiment towards the Liberal party due to his misleading comments regarding covid-19 vaccines. Another potential source of negative sentiment could be the Reserve Bank of Australia increasing the interest rate, leading to an increase in living expenses. Several other issues could also be attributed to the observed increase in negative sentiment, and as polling has shown, the Liberal party is significantly behind the Labor party.

To gain more recent insights on public sentiment the project team developed a live feed of sentiment towards both major political parties and their respective leaders. This visualization showed that both political parties are much closer to each other in sentiment as the election draws closer, suggesting a public opinion swing from Labor toward Liberal as the election draws closer. However, when comparing the sentiment towards both leaders, Albo is clearly the preferred candidate, scoring much higher on total sentiment count and on all positive sentiments. Based on all the data and visualizations throughout the report Labor looks to be the most likely winner of the election. This is further supported by all major polls conducted in 2022.

# CONCLUSION

Ultimately, this research paper aimed to collate data collected from Twitter to determine public opinion and sentiment on Labor and Liberal parties and predict a projected outcome for the 2022 Australian election. Based on the data collected via twarc2, the sentiment analysis observed that both parties shared more negative reactions than positive ones.

Specifically, it was revealed that the public shared more of a positive and trusting outcome with Anthony Albanese and the Labor party. Albanese and the Labor party are significantly more active on Twitter, and the data is slightly skewed towards them; however, once the sentiment analysis was complete, Albanese received quite a higher number of shares and likes on his tweets than his opposition.

It should be noted that several limitations occurred during the research. These include the comparison of the word clouds is that it is not immediately possible to see how keywords have changed over time; they must be analysed and attempted to notice if words are present in one dataset and not in the other. Another limitation is that while we can observe that many of the keywords have emotions associated with them, sentiment towards a party was not clear. To accurately represent the data, the data was cleaned and compiled. The result of that was that Labor and Albanese received more of a trusting and positive outcome than their opposition. **The prediction suggests that the Labor party is likely to be projected to win the 2022 election.**

# REFERENCES

Ramteke, J, Shah, S, Godhia, D & Shaikh, A 2016, *'Election result prediction using Twitter sentiment analysis'*, in 2016 international conference on inventive computation technologies (ICICT), vol. 1, pp. 1-5.

Tumasjan, A, Sprenger, TO, Sandner, PG & Welpe, IM 2011, *'Election Forecasts With Twitter:How 140 Characters Reflect the Political Landscape'*, Social Science Computer Review, vol. 29, no. 4, pp. 402-18.

Bansal B. and Srivastava, S., 2018. *On predicting elections with hybrid topic based sentiment analysis of tweets.* Procedia Computer Science, 135, pp.346-353.

Daniel, W.T. and Obholzer, L., 2020. *Reaching out to the voter? Campaigning on Twitter during the 2019 European elections.* Research & politics, 7(2), p.2053168020917256.

Parmelee, J.H. and Bichard, S.L., 2011. *Politics and the Twitter revolution: How tweets influence the relationship between political leaders and the public.* Lexington books.