Do contributors to NZ political discussion (#nzpol) on Twitter exist in echo chambers?

Social Network Analysis | May 2022

George Moon

Background

New Zealand (NZ) is a developed nation of ~5 million people, governed as a parliamentary democracy. Since 1996, national elections have been run on a 'mixed member proportional' basis, and have always resulted in parliamentary representation from at least five parties. Accordingly, political debate encompasses a wide range of views and political figures, and having a relatively small population means that discussion is mostly conducted at a national level. Twitter is home to many of these debates, under the hashtag '#nzpol.' The community of people who tweet with this hashtag has been described in a media article by 'The Spinoff' as "small and attentive," and "insular"—arguably a reflection of the wider political culture (Mathias 2021). The same article suggests that the same names "[come] up over and over again," and that a significant proportion of active NZ political Twitter users are journalists and politicians themselves.

Research Questions

This research seeks to understand more about the contributors to #nzpol discussion on Twitter through understanding network analysis on the links between users. Two research questions have been defined to guide the enquiry:

1. Do 'echo chambers' exist (or not)? [RQ1]
2. Are there key actors that link between these clusters (if observed)? [RQ2]

'Echo chambers' refers to the "formation of groups of like-minded users framing and reinforcing a shared narrative," and is a phenomenon that can be observed across different social media platforms (Cinelli et al. 2021). In this context RQ1 seeks to understand if there are clusters of users within the network whose grouping could potentially explained by political views.

Assuming that echo chambers are observed, through brokerage analysis RQ2 seeks to understand whether there are key accounts that join these clusters—that is, determining actors that connect otherwise unconnected actors and/or fill gaps in the network (Gould and Fernandez 1989). My hypothesis is that these are the accounts of journalists, news/media organisations or other reputable sources of information—by international standards, New Zealanders' trust in news is still relatively high (Myllylahti and Treadwell 2022).

Methods and Results

The methodology consisted of obtaining all tweets using the hashtag #nzpol for a given period of time, obtaining the 'friend' networks between users using the hashtag, then running network analyses on the results.

```r
library('tidyverse')
library('ggplot2')
library('rtweet')
library('igraph')
library('intergraph')
library('sna')

# API variables stored in .Renviron file
token = create_token(
  app = Sys.getenv("APP_NAME"),
  consumer_key = Sys.getenv("API_KEY"),
  consumer_secret = Sys.getenv("API_KEY_SECRET"),
  access_token = Sys.getenv("ACCESS_TOKEN"),
  access_secret = Sys.getenv("ACCESS_TOKEN_SECRET"),
  set_renv = TRUE)
```

Obtaining and processing tweet data

To build a comprehensive #nzpol dataset, a script was configured on a remote server to retrieve and save #nzpol tweets daily. The Rda files could subsequently be downloaded for processing.

```r
# Get a list of the Rda files
wd <- getwd()
files <- list.files(paste0(wd, '/dataset_cron'))

# Consolidate into a single list
datasets <- lapply(files, function(filename) {
  load(file=paste0(wd, '/dataset_cron/', filename))
  return (tweets_nzpol)
})

# Process datasets into one table, removing duplicates (by checking 'status_id')
tweets <- distinct(data.table::rbindlist(datasets), status_id, .keep_all = TRUE)
```

From here it was a case of filtering the tweets to a three-week period, chosen because of the practical constraints of this exercise.

```r
tweets <- tweets %>% filter(
  created_at >= as.POSIXct("2022-03-14 00:00:00", tz="UTC"),
  created_at <= as.POSIXct("2022-04-03 23:59:59", tz="UTC")
)
```

To get an idea of the nature of the tweets being analyses, a few charts/statistics were generated.
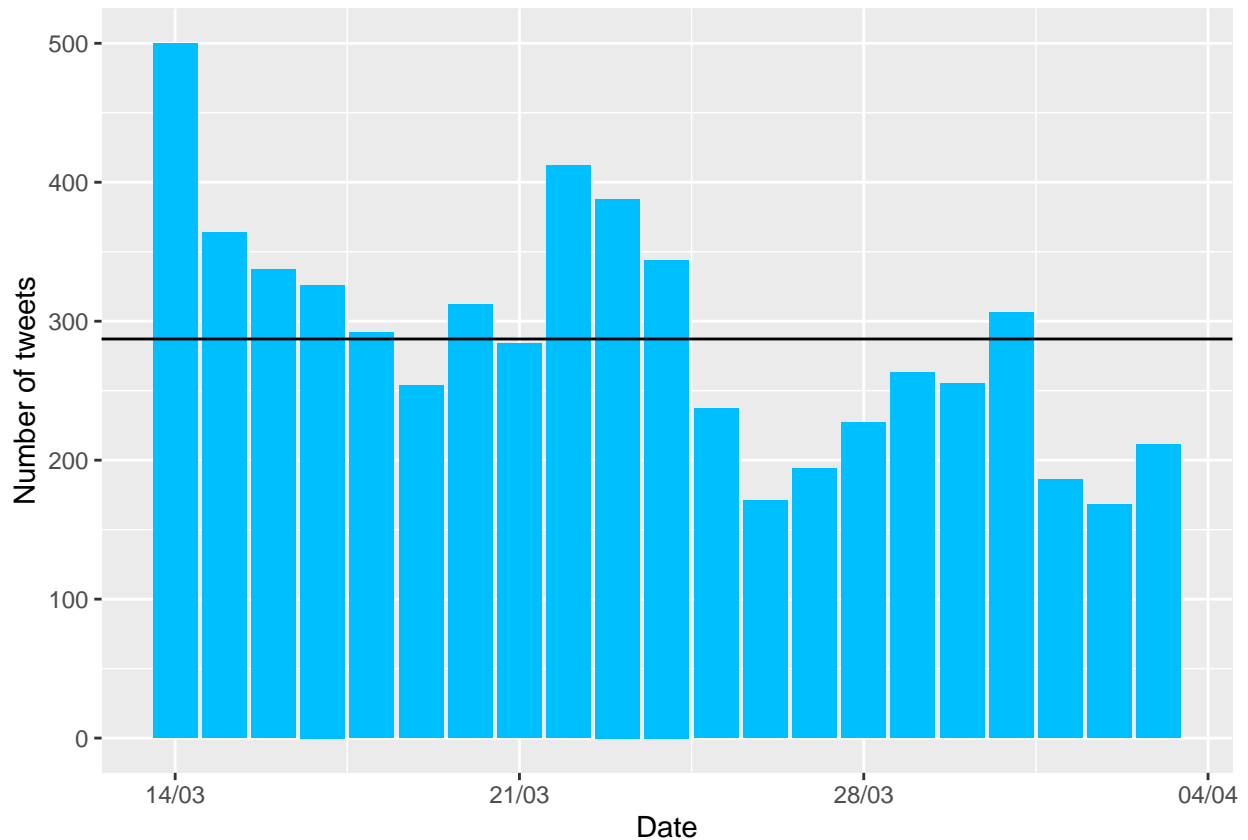
Tweets per day

```r
tweets_by_date <- tweets %>% split(as.Date(tweets$created_at))

tweets_by_date_count <- stack(lapply(tweets_by_date, function(x) {
  as.integer(count(x))
```

```
}))

ggplot(tweets_by_date_count, aes(as.Date(ind), values)) +
  geom_col(fill = "deepskyblue") +
  geom_hline(yintercept=mean(tweets_by_date_count$values)) + # Mean
  scale_x_date(date_breaks = "1 week", date_labels =  "%d/%m") +
  xlab('Date') +
  ylab('Number of tweets')
```
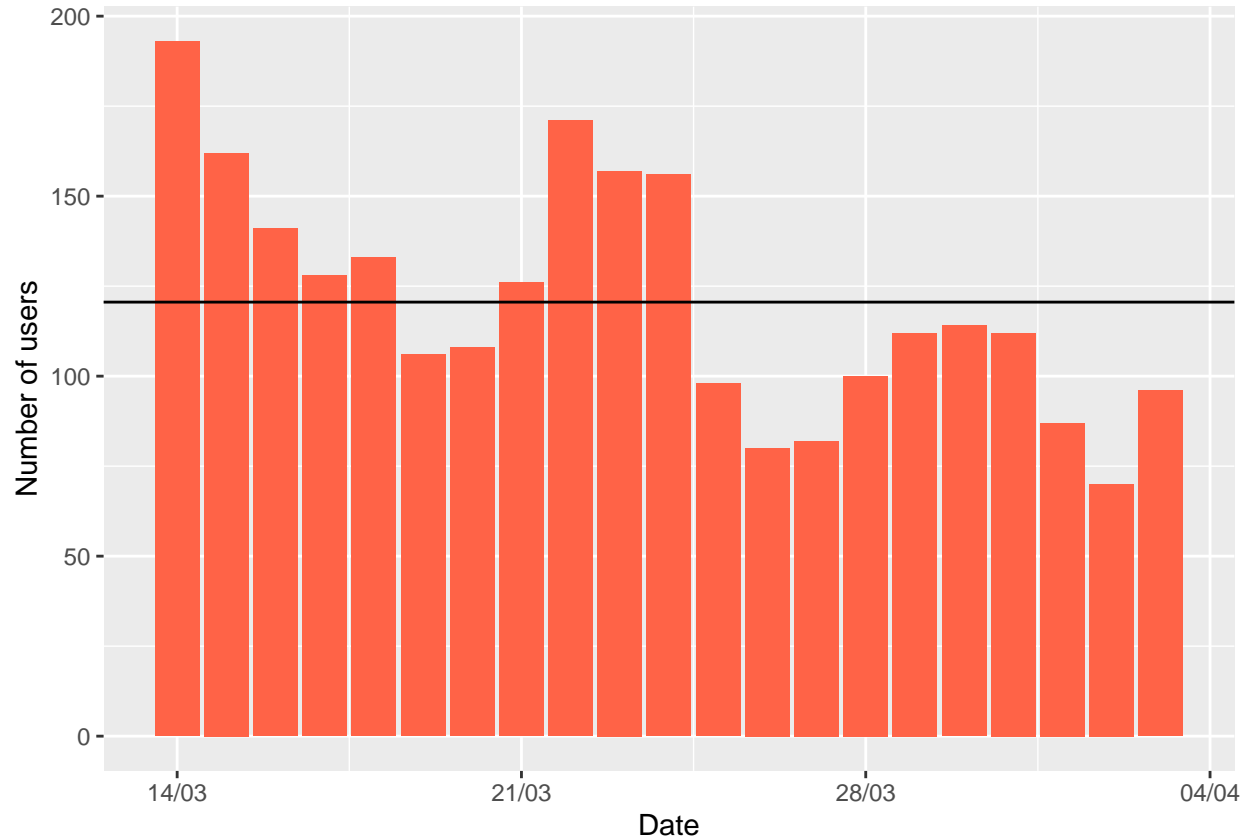

Distinct users per day

```
tweets_by_date_users_count <- stack(lapply(tweets_by_date, function(x) {
  users <- distinct(x[,c('user_id')])
  as.integer(count(users))
}))

ggplot(tweets_by_date_users_count, aes(as.Date(ind), values)) +
  geom_col(fill = "tomato") +
  geom_hline(yintercept=mean(tweets_by_date_users_count$values)) + # Mean
  scale_x_date(date_breaks = "1 week", date_labels =  "%d/%m") +
  xlab('Date') +
  ylab('Number of users')
```

Data statistics

```
count(tweets) # Total number of tweets
```

```
##       n
## 1: 6031
```

```
count(distinct(tweets[,c('user_id')])) # Number of distinct users
```

```
##      n
## 1: 813
```

Visually it appears that distinct users and the quantity of tweets are quite similar, suggesting that the number of tweets is related to the number users engaging in the conversation, rather than a few users tweeting repeatedly.

Obtaining and processing friend networks

Using the list of tweets, a list of user accounts were obtained and validated to ensure that they were neither suspended nor protected.

```r
# Get list of users
users_from_tweets <- tweets[,c('user_id','screen_name', 'friends_count')] %>%
  distinct(user_id, .keep_all = TRUE)

# Only include those with at least one friend
users_from_tweets <- users_from_tweets[
  which(users_from_tweets$friends_count > 0),
]

# Obtain updated user account information
users <- users_data(
  lookup_users(users_from_tweets$user_id, token = token, parse = TRUE)
)

# Filter users by valid, non-protected accounts
users <- users %>% filter(
  users$user_id %in% users_from_tweets$user_id,
  users$user_id %in% filter(users, users$protected == FALSE)$user_id
)

# Rename 'name' field to prevent undesired automatic use by 'sna' analyses
users <- users %>% rename(twitter_name = name)

count(users)
```

```
## # A tibble: 1 x 1
##       n
##   <int>
## 1   776
```

With this user list, it was then possible to collect data on the links between accounts ('friends'). Due to API constraints, this is a time-intensive operation thus the (previously collected) dataset will be loaded. This analysis focuses on links between users related to #nzpol, thus we will filter out accounts not relevant to this.

```r
#friends <- get_friends(users$user_id, retryonratelimit = TRUE, token = token)
load('./final_dataset/friends.Rda')

# Filter irrelevant accounts
friends <- friends %>% filter(
  friends$user %in% users$user_id,
  friends$user_id %in% users$user_id,
  friends$user_id %in% friends$user
)
```

Adding political leanings

To understand more about the presence (or absence) of echo chambers, a classification system was determined for users. Accounts were labelled either as 'news' accounts, representing journalists or neutral institutions (such as parliament); or they were labelled as 'left' or 'right,' representing political leanings.

This classification was completed manually by reading Twitter account biographies, looking for key terms. This simple approach has been used elsewhere for classifying political leanings (Ramaciotti Morales, Cointet,

and Zolotoochin 2021). In this instance a certain amount of subjectivity was necessary, but accounts were classified only if they demonstrated a clear leaning, or were political commentators with well-known views.

Example of classification terms used:

- **Left:** 'leftie,' 'ardern supporter,' 'labour'

- **Right:** 'conservative,' 'individual,' 'freedom,' 'liberty,' 'national'

- **News:** organisations ('TVNZ,' 'RNZ,' 'Discovery,' 'Stuff,' etc.), official emails, 'journalist'

In some instances it was necessary to carefully read biographies to ensure the user was not expressing opposition to an ideology (e.g. "[I] like arguing with lefties"). An equal number of accounts were classified as left/right.

```r
# Export user details (as-is)
#write.csv(users, './final_dataset/user_details.csv')

# Import (manually-processed) leanings
leanings <- read.csv(
  './final_dataset/political_leanings.csv',
  stringsAsFactors = TRUE,
  na.string = "", # Empty strings should be NA
  colClasses=c("user_id"="character") # Treat numeric user_id as characters
)
leanings <- select(leanings, -screen_name) # Ignore 'screen_name' (if present)
users <- users %>% left_join(leanings, by = c("user_id" = "user_id"))

# Show breakdown of attributed 'leaning' values
users %>% count(leaning)
```

```
## # A tibble: 4 x 2
##   leaning     n
##   <fct>   <int>
## 1 left       25
## 2 news       15
## 3 right      25
## 4 <NA>      711
```

Adding style attributes

Colours, shapes and sizes were added to the dataset to facilitate visual understanding of the network graphs.

```r
# Make identified accounts bigger, and 'news' accounts the largest
users <- users %>% mutate(size =
  case_when(
    .$leaning == 'left' ~ 4,
    .$leaning == 'right' ~ 4,
    .$leaning == 'news' ~ 7,
    TRUE ~ 2
  )
)
```

```
# Assign colours (used for basic plot; ignored by clustering plot)
users <- users %>% mutate(color =
  case_when(
    .$leaning == 'left' ~ 'red',
    .$leaning == 'right' ~ 'blue',
    .$leaning == 'news' ~ 'black',
    TRUE ~ 'grey'
  )
)

# Assign shapes (to help differentiate account types for clustering plot)
users <- users %>% mutate(shape =
  case_when(
    .$leaning == 'left' ~ 'circle',
    .$leaning == 'right' ~ 'square',
    .$leaning == 'news' ~ 'circle',
    TRUE ~ 'circle'
  )
)
```

Generating the graph

From here it was possible to generate the network graph. As a simple means of removing 'unimportant' accounts from the graph—those unlikely to be of value to the analysis—accounts with two or fewer degrees (of connectivity) were filtered.

```
graph <- graph_from_data_frame(friends, directed = FALSE, vertices = users)
graph <- delete_vertices(graph, igraph::degree(graph) <= 2)
```
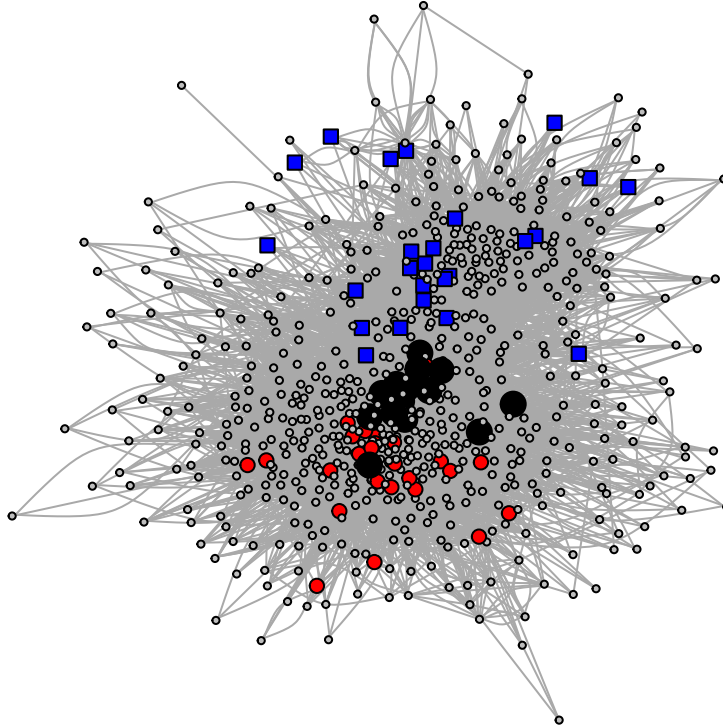
The graph was then able to be plotted, with labelling disabled to facilitate visualisation.

```
plot(graph, vertex.label=NA)
```

Visually it is immediately apparent that the 'left' accounts (medium red circles) and 'right' accounts (medium blue squares) are generally clustered on opposing sides of the graph, with 'news' accounts (large black circles) positioned reasonably centrally. This provides evidence in support of echo chambers [RQ1] and 'news' accounts acting as brokers [RQ2], however this should be confirmed by qualitative analyses.

Network analyses

Community identification

Using the 'walktrap' algorithm, it is possible to identify communities within the graph, that is, parts of the graph that are "highly connected between them but with few links to other vertices" (Pons and Latapy 2005). If echo chambers exist, at least two communities would be expected and the account membership (accounts that make up each community) would be expected to be stratified along political leaning divides.

```
communities <- cluster_walktrap(graph)

plot(communities, graph, vertex.label=NA, edge.color = 'gray')
```
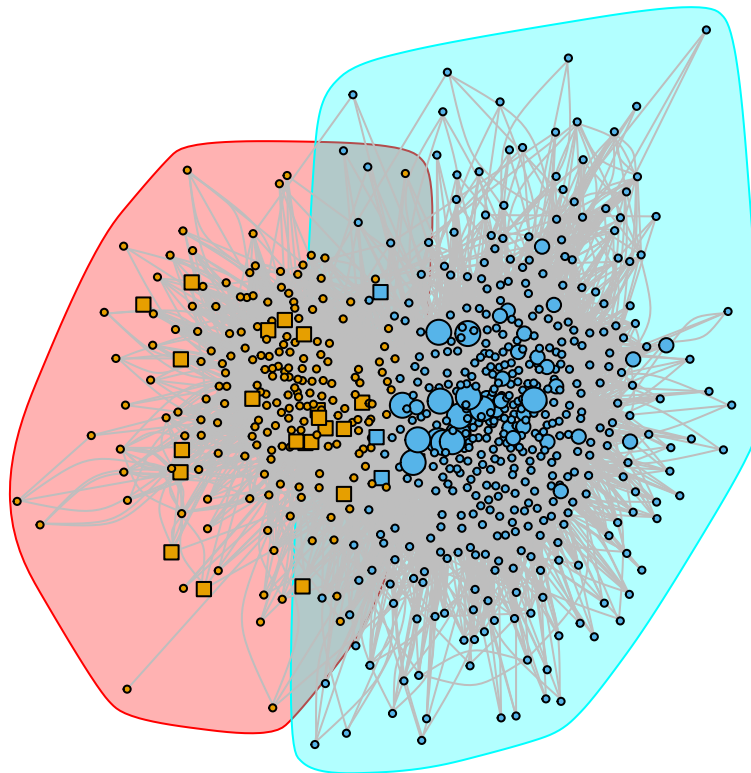
Indeed, two communities are present and it appears that all of the 'left' accounts (medium circles) are members of one community, and almost all of the 'right' accounts (medium squares) are members of the other community. 'News' accounts (big circles) all appear to be a part of one community.

This can, of course, be verified by numbers:

```
users_membership <- data.frame(communities$names, communities$membership)

users <- left_join(
  users, users_membership, by = c("user_id" = "communities.names")
)

filter(users, !is.na(users$leaning)) %>%
  group_by(communities.membership, leaning) %>%
  summarise(count = n(), .groups='keep')
```

```
## # A tibble: 4 x 3
## # Groups:   communities.membership, leaning [4]
##   communities.membership leaning count
##                    <dbl> <fct>   <int>
## 1                      1 right      21
## 2                      2 left       25
## 3                      2 news       15
## 4                      2 right       4
```

The numbers confirm the visual analysis, and provide strong evidence for the 'echo chamber' effect [RQ1], with 'left'/'right' accounts being almost completely stratified along political lines.

Identifying brokers

To verify the presence (or absence) of brokers [RQ2], brokerage scores were calculated using the Gould-Fernandez brokerage analysis (Gould and Fernandez 1989). From here a histogram could be generated to get an understanding of the distribution of brokerage values across accounts.

```r
network <- asNetwork(graph)

brokers <- brokerage(network, cl=get.vertex.attribute(network, 'user'))

broker_scores <- as.data.frame(brokers$raw.nli)
broker_scores$user_id <- row.names(broker_scores)

# Append brokerage scores to users
users <- left_join(users, broker_scores, by = 'user_id')

# Plot histogram (y-axis limited on purpose to facilitate visualisation)
# hist(users$t, breaks=100, ylim=c(0, 80))

# Plot histogram
hist(users$t, breaks=100)
```
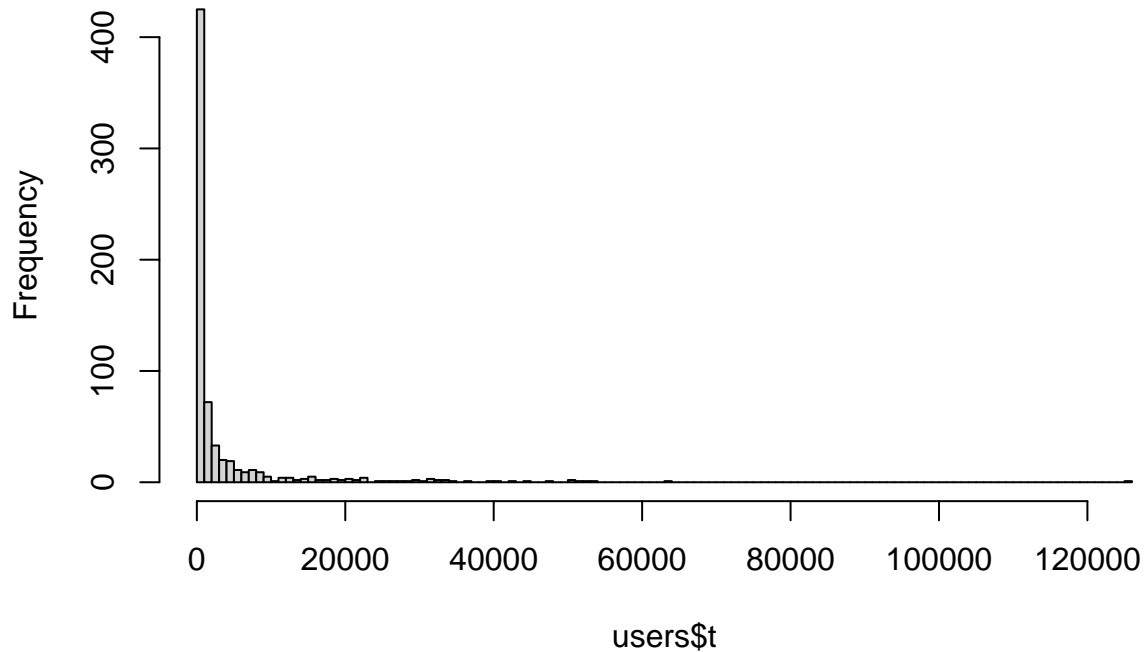
Histogram of users$t

As could be expected, the vast majority of accounts had low brokerage values, but there were a few accounts with comparatively high values. To test the theory that 'news' accounts are key brokers, they were differentiated from 'non-news' accounts and then brokerage values graphed on a box plot. If 'news' accounts were key brokers, then it would be expected to see higher values in the box-plot for the 'news' accounts.

```r
users <- users %>% mutate(isNews = if_else(.$leaning == 'news', TRUE, FALSE))

# Numeric
# users %>% group_by(isNews) %>% summarize(mean = mean(t), median = median(t))

# Box plot
ggplot(filter(users, !is.na(leaning)), aes(x=t)) +
  geom_boxplot() +
  facet_grid('isNews', labeller=label_both) +
  ggtitle('News accounts vs. non-news accounts')
```

News accounts vs. non–news accounts
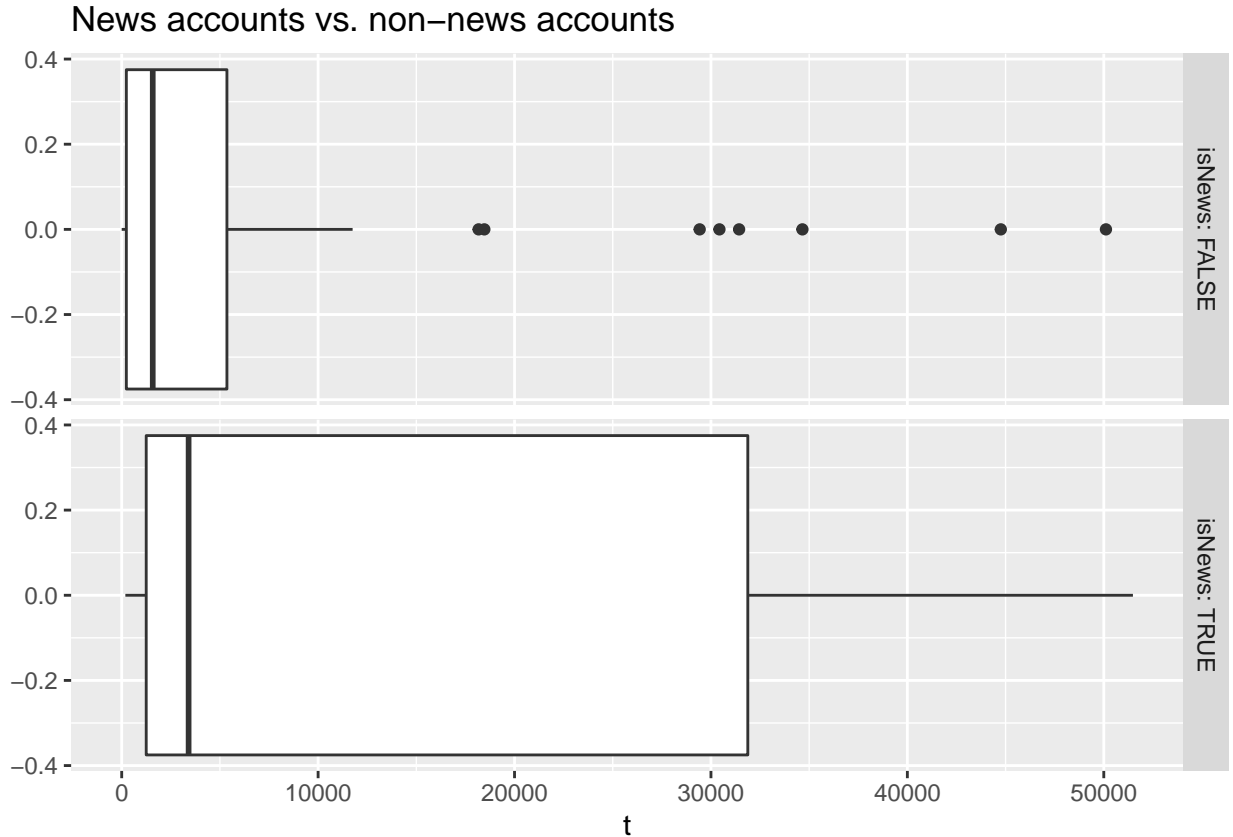
Indeed, the box plots provide evidence that 'news' accounts act as key brokers in the network—the median value for 'news' accounts is greater than that of 'non-news' accounts. Whilst there are a few 'non-news' accounts that have high brokerage scores, these are statistical outliers.

Discussion

The analyses suggest that there are echo chambers, divided along political leanings [RQ1] and key brokers, which are often 'news' accounts [RQ2]. Nonetheless, there are a number of limitations to this research.

Regarding data collection and processing, the #nzpol hashtag only collects a subset of Twitter discussion, and data collection was time-limited. A more complete analysis might include all accounts linked to NZ politicians/political actors, and be conducted over a longer period. In addition, 'political leaning' identification was simplistic and somewhat subjective, limitations which could be overcome by conducting algorithmic sentiment analysis of accounts, or introducing more classification categories (Preoţiuc-Pietro et al. 2017; Ramaciotti Morales, Cointet, and Zolotoochin 2021).

In the network analyses, the walktrap function was used with the default steps number (4). If this value is altered, it outputs a different number of detected communities, and a more detailed analysis should consider the implications of changing these values. For example, the potential reasons why the 'news' accounts were grouped in the same community as the 'left' accounts was not analysed, but further exploration with the walktrap function might provide some clues.

Similarly, the brokerage function only tells us that certain accounts are key to 'filling gaps' in the network, but more detailed analysis might provide information on the nature of these links, as in, whether they are actually linking the echo chambers, or they are linking disparate accounts within them. In addition, not all news organisations can be considered as 'neutral,' so the potential biases of these accounts could be included.

Despite these limitations, this analysis is still a useful insight into NZ political debate. Whilst it is hard to say how much Twitter discussion has implications for wider society, politicians and journalists do actively use it as an avenue for sharing and engage with the public—and there have been examples of social media pressure causing government action (Culture and Heritage 2016; Mathias 2021). In addition, an increasing number of people get their news from social media, thus understanding more about these networks can help with tracking the spread of information, and combatting disinformation (Myllylahti and Treadwell 2022).

Conclusion

To conclude, in order to understand more about the dynamics of NZ Twitter political discussion, an analysis of accounts using the #nzpol hashtag was conducted. This was done using visual plots, community detection and brokerage analysis, following the classification of accounts' by political leanings or 'news' status. The analyses provided evidence of echo chambers along political lines, and also the role of 'news' accounts acting as key brokers in the network. Future research could address some of the limitations of this study, for example by looking at a wider group of Twitter accounts or introducing different processes of classification. Whilst it is difficult to fully understand the wider societal implications, network analysis like this still provides useful insight into political debate, and understanding more about social networks can help with tracking information and disinformation.

References

Cinelli, Matteo, Gianmarco De Francisci Morales, Alessandro Galeazzi, Walter Quattrociocchi, and Michele Starnini. 2021. "The Echo Chamber Effect on Social Media." *Proceedings of the National Academy of Sciences* 118 (9): e2023301118. https://doi.org/10.1073/pnas.2023301118.

Culture, Ministry for, and Heritage. 2016. "Red Peak Flag." https://nzhistory.govt.nz/media/photo/red-peak.

Gould, Roger V., and Roberto M. Fernandez. 1989. "Structures of Mediation: A Formal Approach to Brokerage in Transaction Networks." *Sociological Methodology* 19: 89–126. https://doi.org/10.2307/270949.

Mathias, Shanti. 2021. "Does #Nzpol Actually Impact Politics?" https://thespinoff.co.nz/irl/18-11-2021/does-nzpol-actually-impact-politics.

Myllylahti, Merja, and Greg Treadwell. 2022. "In Media We Trust? A Comparative Analysis of News Trust in New Zealand and Other Western Media Markets." *Kōtuitui: New Zealand Journal of Social Sciences Online* 17 (1): 90–100. https://doi.org/10.1080/1177083X.2021.1948873.

Pons, Pascal, and Matthieu Latapy. 2005. "Computing Communities in Large Networks Using Random Walks (Long Version)." *arXiv:physics/0512106*, December. http://arxiv.org/abs/physics/0512106.

Preoţiuc-Pietro, Daniel, Ye Liu, Daniel Hopkins, and Lyle Ungar. 2017. "ACL 2017." In, 729740. Vancouver, Canada: Association for Computational Linguistics. https://doi.org/10.18653/v1/P17-1068.

Ramaciotti Morales, Pedro, Jean-Philippe Cointet, and Gabriel Muñoz Zolotoochin. 2021. "Unfolding the Dimensionality Structure of Social Networks in Ideological Embeddings." In. Amsterdam, Netherlands. https://hal.archives-ouvertes.fr/hal-03315759.