"Power to the People!": Social Media Discourse on Regional Energy Issues in Australia

Kerri Morgan
Monash University & Deakin University
KerriMorgan@deakin.edu.au

Marc Cheong
Monash University

Susan Bedingfield
Monash University

Abstract

Social media provides people from all socio-economic sectors with the opportunity to voice their opinions. Platforms such as Twitter provide the means to share one's opinion with little effort and cost. But do these media empower everyday people to make their voice heard? In this research, we introduce a novel approach for investigating the voice of different Twitter groups on social media platforms, by combining text clustering and an analysis of cliques in the resulting network. We focus on a case study using Twitter interactions with respect to energy issues, in particular the closure of coal-fired power stations such as Hazelwood. Implications from this study will benefit stakeholders from governments to industry to the 'common man', in understanding how discourse on social media reflects public consumer sentiment.

Keywords: social media, Twitter, graph theory, data mining, pattern recognition.

1 Introduction

Recently, there has been an increasing global focus on green energies (Byrnes, Brown, Foster, & Wagner, 2013; Furlan & Mortarino, 2017; Twidell & Weir, 2015). Some consequences of this are an increasing number of new green technologies (Bilgili, Ozbek, Sahin, & Kahraman, 2015; Kolokotsa, 2017; Sen & Bhattacharyya, 2014), the increasing likelihood of energy brownouts and blackouts (Puddy, 2015), and the closure of traditional coal fired power plants (Davies, 2014; Evans & Phelan, 2016; Johnstone & Hielscher, 2017).

While in the past, wide-reaching discussion of issues of national or global importance has been in the domain of the press, industry leaders and politicians, the rise of social media has provided the opportunity for other stakeholders, in particular the *vox populi* of the 'common man', to voice their opinions on similar issues (Shamma, Kennedy, & Churchill, 2009; Dudley-Nicholson, 2016).

In this paper, we consider the power of the voice of the 'common man' in social media. We present the results of our pilot study using the discussion on Twitter about the closing of Hazelwood power station, an event that affects a number of social issues such as employability and thus could be expected to trigger comments from a wide range of people.

Clustered behaviour is important in this context as groups of people often function as cohorts on social media. For this reason, we cluster the tweet contents in order to identify possible cohorts of users with common opinions. We then analyse the interactions between these

clusters using a new approach utilising clique analysis on the clustered tweet contents. The metadata for the Twitter users in each cluster is illustrated using Tableau.

2 Background

2.1 Case Study: The Hazelwood Power Station

The discussion surrounding the closure of the Hazelwood power station in the La Trobe Valley, Australia, has provoked a variety of issues. These include the use of green energies, the cost of energy, the sufficiency (or insufficiency) of existing energy sources, and the uncertainty of employment for the Hazelwood employees, and other related social issues (Anderson, 2017).

In this paper, we examine the interactions between different Twitter stakeholders in the Australian context, specifically around Hazelwood, by studying the following key themes:

interaction patterns, the impact of the views of the 'common man' and the behaviour of distinct groups of Twitter users.

2.2 Twitter

Dating back to initial experimental studies in 2007-2009, Twitter data has been used extensively across academic disciplines to study human phenomena – manifest in discourse patterns and group behaviours in social media – in areas ranging from election campaigns to natural disasters to celebrity worship (Cheong, 2013).

We define several key terms as follows. *Tweet* refers to a message sent on the Twitter microblogging service, which is at most 140 characters long. We use the term tweet and message interchangeably. A *Tweeter* is a colloquialism for a Twitter user. Both terms - user and tweeter - will be used interchangeably. A *#hashtag* is a keyword preceded by a hash (#) symbol that allows users to tag their tweets as belonging to a certain topic. A *@user* is a kind of message that is directed at, or replies to, another user, akin to an email reply. Finally, a *retweet*, or an *RT*, indicates messages that are re-broadcasted by another user, similar to the concept of forwarding an email message. A full literature review of the general properties of Twitter can be found in (Cheong, 2013).

Twitter has an Application Programming Interface (API) which allows researchers to download a proportion of tweets in an automated fashion. In addition to the tweets themselves, additional metadata about the tweet (e.g. timestamp, device used) as well as the users associated with the tweet (Twitter user name, profile picture, etc.) can be collected to enrich data analysis.

2.3 Clustering and Pattern Recognition

In this study, we identify distinct cohorts of users based on the textual content of their tweets.

Clustering is a common method of segmenting structured data into similar subgroups, and is often used in social media research focusing on groups (e.g. Zhang, et al., 2007 and Jansen, et al., 2011). Determining which data objects are similar is based on a predefined metric relevant to the particular data set. An overview of clustering algorithms is available in (Xu & Wunsch, 2005) and (Abbas, 2008). Specific text clustering algorithms can be applied to textual data such as a tweet. A useful survey of the currently used text algorithms can be found in (Aggarwal & Zhai, 2012).

2.4   Representing a Social Network as a Graph

A *graph* is a set of *vertices* (or nodes) that model objects and *edges*, connecting pairs of vertices, that model relationships between the objects.   Graphs are used to model a wide range of networks. Examples include environmental management (Aittokallio & Schwikowski, 2006; Bunn, Urban, & Keitt, 2000), biological networks (Mason & Verwoerd, 2007), social networks (Freeman, 1977) and electricity networks (Holmgren, 2006). In this paper, we use standard definitions for graph concepts: see for example (Bang-Jensen & Gutin, 2008; Diestel, 2005).   A *clique* is a subset of vertices in which every vertex is connected to every other vertex in the set by an edge. A (connected) *component* in a graph is a maximum set of vertices such that every pair of vertices in the set are connected by a path of edges.

Graphs are invaluable tools in modelling social networks, dating as far back as the 1930s (Moreno, 1934). An example of a graph representation of a real-world social network is illustrated in Figure 1 (Bearman, Moody, & Stovel, 2004).
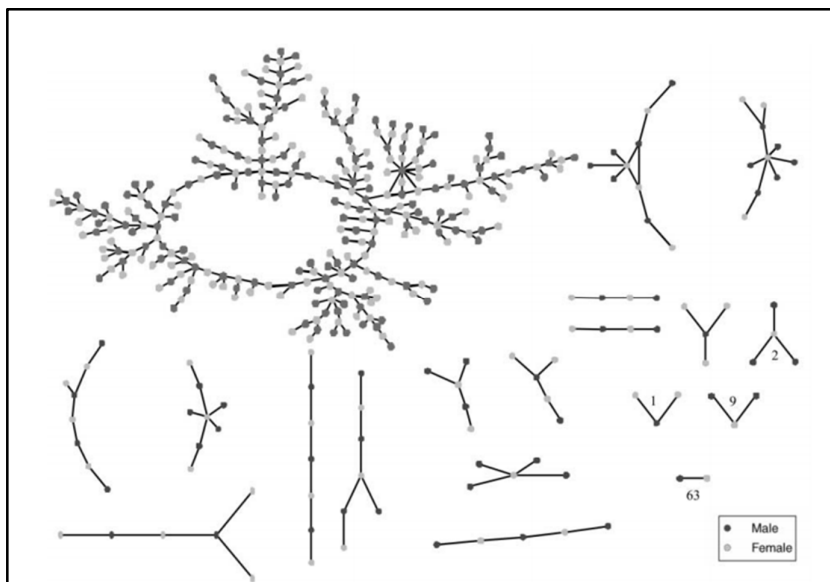

Figure 1. Graph visualisation of a real-world social network. This is meant to represent romantic relationships between students in a high school in the USA. Figure reproduced from Bearman, et al (2004).

The use of graphs in social network modelling and representation is even more relevant in the analysis of online social networks (OSNs) due to the richness of information that OSNs provide. In this paper, we use standard definitions and concepts found in social network literature; see for example (Wasserman & Faust, 1994).

We designed our methodology around the research themes presented in Section 3. Based on the structure of the Twitter social network and its general characteristics, four research questions are proposed.

3   Research Questions

RQ1. Do the communication and engagement patterns reflect those of a typical social network?

A typical online social network, such as Twitter, exhibits 'scale-free' behaviour (Barabási and Albert, 1999), in contrast to localised real-world social networks, which are characterised as 'small world' (Cheong, 2013; Green, 2014). Online social network users can form connections with people at any geographic location (Green, 2014). An important characteristic of a scale-free network is that popular people "acquire hundreds, thousands, or even millions of [online] 'friends'", leading to a hierarchical distribution (Green, 2014).

In RQ1, we examine the layout of communication and engagement between Twitter users (and indirectly, messages) discussing the Hazelwood power station issue. In particular, we examine the chain of @user replies to illustrate the flow of information between users.

Is the discourse between users about Hazelwood typical of the scale-free behaviours exhibited on Twitter and OSNs in general?

RQ2. Does clustering effectively identify user groups?

As described in Section 4.2, the tweets in our sample are grouped into distinct clusters so that tweets with similar content are grouped together. Each tweet cluster has a group of users associated with it. For convenience, we say the user 'belongs to' the cluster. A given user may belong to more than one tweet cluster depending on the range of topics that they tweet about. The number of clusters a given user belongs to ranges between 1 and 30 with the majority, approximately 67%, of users belonging to exactly one cluster and approximately 15% of users belonging to exactly two clusters. The remaining 18% minority of users belong to multiple tweet clusters implying that these tweeters used Twitter to communicate about multiple issues during the sampling period. We then analyse the users belonging to each tweet cluster to determine whether there are specific user characteristics associated with each cluster's tweet group.

RQ3. Using graph-theoretic approaches, can we identify emergent communication patterns between clusters of Twitter users?

We examine the communication flow between the groups of users associated with each cluster. Two groups of users associated with two different clusters are connected if there exists a directed path connecting the two clusters in the model of the network. The minimum distance of the path gives an indication of how closely the two groups of users are connected. The identification of groups of users that only communicate with other users within their own cluster may indicate that these users are essentially powerless in the network with respect to the tweet contents related to that specific cluster.

RQ4. Does the 'common man' have a 'powerful' voice on social media? More explicitly: Do Australian Twitter users 'amplify' their own voice, the voice of politicians, the media, and opinion-leaders (Retweets)?

A central theme in this research is to determine whether there is a group, or groups, of users that can be identified with 'the common man' and the effectiveness of their voice in communicating with other users in the network.

4 Methodology and Results

4.1 Twitter Data Collection

4.1.1 Collection Method

Twitter metadata was collected via the Twitter API using a PHP data collection framework. (Borra & Rieder, 2014)

The Twitter network was polled based on the following search queries.

> *@Greens* - the official Twitter username of the Australian Greens, an environmentally-conscious political party that campaigned for Hazelwood's closure[1]

> '*hazelwood power*' - a wildcard matching references to the Hazelwood power station

> '*snowy hydro*' - a wildcard matching references to the Snowy Mountains Hydro-electric Scheme, in order to capture related discourse[2].

Data was collected from 23 March 2017 to 14 June 2017, and includes the date of the official closure of Hazelwood (29 March 2017[3]), as well as during the resulting post-closure aftermath. Due to the limitations of the Twitter API on downloads, approximately 10% of all tweets containing the three search queries were collected (Cheong, 2013; Borra & Rieder, 2014). The tweets in our data collection thus represent 10% of all available tweets, selected at random by the Twitter API.

4.1.2 Properties of the Twitter Corpus

The corpus consists of 43,081 tweets with attached metadata records, originating from 9,075 unique Twitter users. The sample size of tweets and messages are commensurate with – if not better – than traditional studies on social networks' messaging patterns based on "observable communications" (Quintaine & Kleinbaum, 2011; Bernard, Killworth, & Sailer, 1980).

The entire corpus is imported into the Tableau BI suite to visualise the distribution of Twitter messages over time (Figure 2). All times are normalised to UTC (Universal Coordinated Time), to avoid issues with daylight savings time ending in Victoria partway through the data collection window.

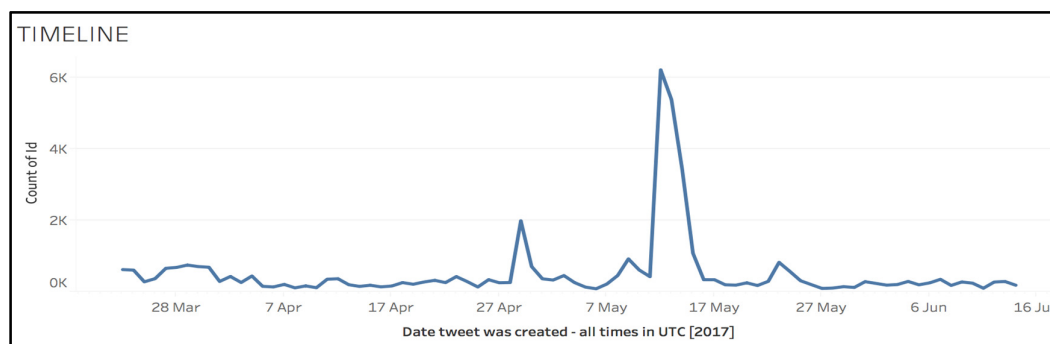*Figure 2. Volume of Twitter messages per day (normalised to UTC), as visualised in Tableau.*

---

[1] An example of their press releases: https://greens.org.au/magazine/vic/wins-hazelwood-transition-long-way-go

[2] See http://www.australia.gov.au/about-australia/australian-story/snowy-mountains-scheme

[3] Press release: http://www.ipplc.com.au/media/newsitem/End-of-generation-at-Hazelwood

The first spike in the data coincides with reactions to Prime Minister Malcolm Turnbull's plan to "underwrite the rail line for… [the] proposed massive Adani central Queensland thermal-coal mine", a controversial issue as the project "threatens their livelihoods" (Hull, 2017). The second spike concerns the Australian Greens supporting the "Turnbull government… [in] overhauling school funding" (Knott, 2017).

The distribution of the dataset can be divided into three basic types of tweets:

1. @user messages - there are 10,100 messages with the prefix "@user", to denote addressivity of a particular tweet to another Twitter user.

2. Retweet or RT messages - there are 25,940 messages with the prefix "RT" to denote retweeting of another's message, either paraphrased or verbatim.

3. "Ether" messages - the remaining 7,041 messages are messages that are not addressed to, or targeted at, anyone in particular - i.e. posting a generic 'status update' for public consumption. We term these non-directed messages "ether messages", as they are akin to broadcasting a message to the *ether*.

These groups will be revisited in Section 4.2 when we apply clustering techniques.

4.1.3 Addressing RQ1

RQ1. Do the communication and engagement patterns reflect those of a typical social network?

In addressing RQ1, we represent the users in our corpus, as well as the connections between them, in terms of a social network graph. We must first define what a *connection* is. Social networks such as Twitter can contain different kinds of connections between users (Cormode, Krishnamurthy, & Willinger, 2010). These include *following* (X follows updates posted by Y), *retweeting* (X is forwarding/retweeting content made by Y), and *replying* (X sends @user messages to Y).

To answer RQ1, we intend to examine the layout of communication and engagement between Twitter users discussing the Hazelwood power station within the timeframe of data collection. Hence, we define a *connection* between users as *replies*, or @user messages, between users[4]. Having defined connection, we now extract all messages from our corpus that only match the @user behaviour described above.

We determine if our corpus contains scale-free behaviour (Cheong, 2013; Green, 2014). A scale-free online social network is typified by a power-law distribution of connections between users. Briefly, a *power-law* is defined as a relationship between two variables "in which a relative change in [the independent variable] ... gives rise to a proportional relative change in the [dependent variable] ... independent of [their]... initial size" (Bar-Yam, 2011). In this case study, our null hypothesis is that the frequency distribution of messages received fits a power law distribution. To test this, we generated a frequency distribution of @user message recipients in Microsoft Excel, as illustrated in Figure 3.

---

[4] 'Following' is more longitudinal in nature and reflects long-term links or similarities between users, while studying 'RT' connections are normally used for studying the 'amplification' of an influencer's message (e.g. politicians, celebrities, newsmakers). Therefore, 'following' and 'RT' links are not suitable for answering RQ1.
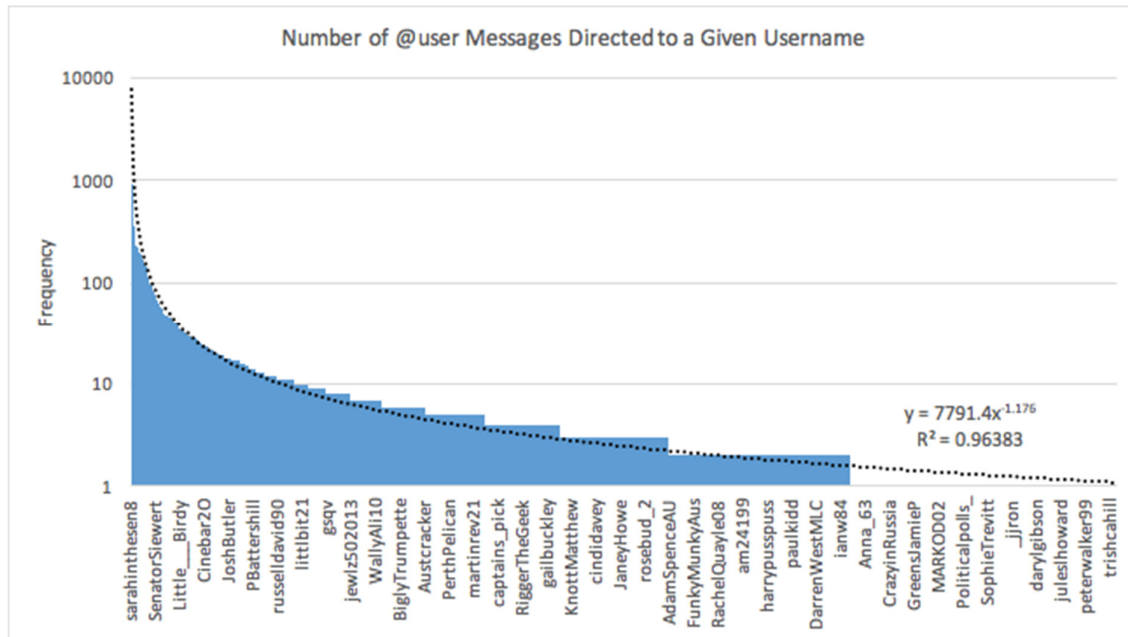
Figure 3. Frequency distribution of @user messages received by a user in our corpus.

A power-law distribution trendline illustrated by the dashed line, was generated for the frequency distribution. A Pearson R-squared coefficient of best-fit of 0.9638 was obtained which indicates a close fit between the empirical distribution and the estimated power-law model.

In our investigation, we also visualise the distribution of such @user connections between users. We employ the Pajek network package (Batagelj & Mrvar, 2004) to preprocess the raw graph data and Gephi (Bastian, Heymann, & Jacomy, 2009) to generate the final visualisation with emphasis on hubs. Figure 4 visualises the network, with directed edges (arcs) representing the presence of connections between users (vertices).
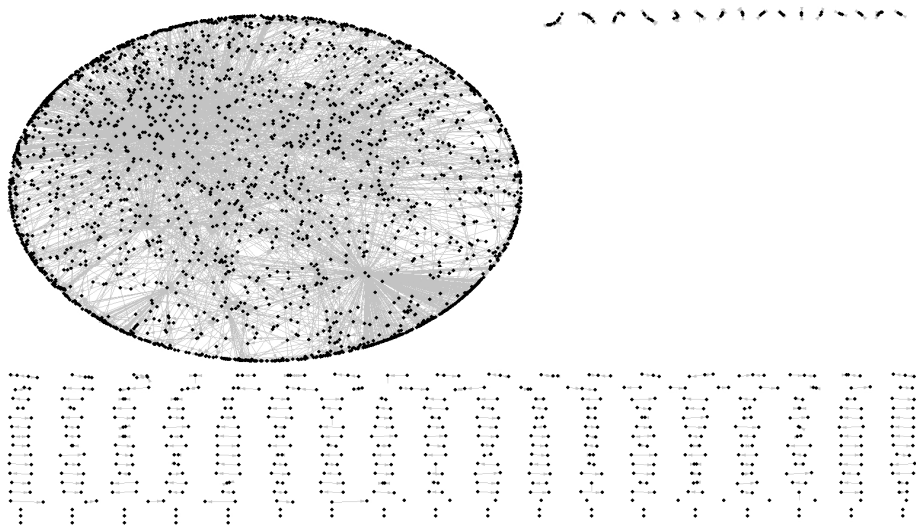
Figure 4a: Initial Pajek network visualisation of all @user connections (grey lines) amongst target users (black dots) found in the corpus.
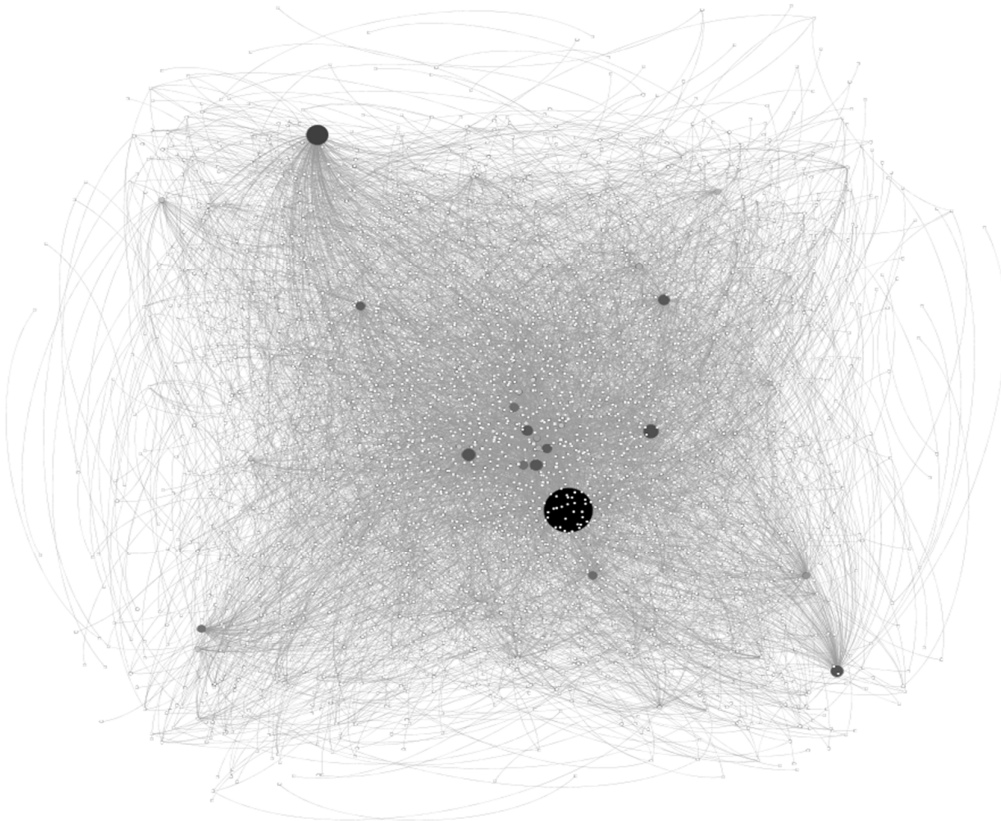
Figure 4b: The same network, stylised in Gephi to emphasise popular 'hub users' (filled nodes) -- circle diameter and colour intensity are proportional to the number of messages (edges) received by a given user.

From Figure 4a, a giant connected component is obtained from the graph visualisation. However, to illustrate the scale-free nature of our network with better clarity, we stylised the graph using Gephi to emphasise the presence of large hubs: a node (user) has size and colour proportional to the number of edges (@user messages) received. Figure 4b illustrates the stylised version, where a few 'hubs', i.e. popular users with many @user messages received, can be clearly seen.

Our results above answer RQ1 in the affirmative - communication patterns relating to our chosen topic, reflected in @user messages between users, does reflect those exhibited in a typical social network. Our investigations also show from that analysis at a macro level, i.e. Figure 4, it is hard to visualise finer-grained detail about constituent users, considering the sheer number of connections. This leads us to conclude that it is useful to organise the users in distinct cohorts. Ergo, we cluster the users by their unique texts, which will be the focus of RQ2 below.

4.2 Clustering of Tweets

For this research, distinct cohorts of tweets were identified by clustering the text contained in the tweets using the K-means algorithm. The content of each tweet was prepared for clustering by implementing established text pre-processing techniques including tokenising, stemming, and filtering tokens and stop words. The processing also included generating two word phrases, or n-grams where n=2. The data was then converted into a term frequency–inverse

document frequency (TF-IDF) matrix (Aggarwal & Zhai, 2012) which is a structured form of the data used in the clustering process. The matrix was pruned so that the frequency of the tokens lay between 3 and 30 percent.

The output of the clustering process was then used to create a social network graph. The structure of this graph was used to analyse discussions based on identified themes in the tweets. This process is described in Section 4.3.

4.2.1 Data Analysis

The processed tweets were clustered and a collection of 40 clusters (cohorts of tweeters) was identified. The clusters identified distinct groups of users based on the content of their tweets. We note that the K-means algorithm requires a specified value for K. In this case a value of K=40 was used as it provided a collection of distinct groups of tweeters that could be further analysed.

A directed graph was then created to represent the network of interactions between these groups of tweeters. After clustering, we also identified key features of users in those clusters including the tweeter's name and description (from the bio-data provided upon sign up to Twitter).

4.2.2 Question RQ2. Does clustering effectively identify user groups?

The clustering is used to identify tweets with similar content. In Section 5.1 we show that individual clusters of interest can be associated with distinct groups of users.

4.3 Network Analysis

In this section, we identify features in the Twitter network by modelling it as a graph. In Section 5, the relationship between the graph and the underlying Twitter network will be discussed. We analysed the network using the following graph properties: the largest/smallest clique to identify groups of users and individual users, source and sink vertices, that is, vertices with no in-going or no outgoing edges respectively, and the shortest directed path or trail between pairs of vertices to indicate the possible extent of individual cluster's influence.

Two types of graphs were used to model the Twitter network. The first was a directed graph that was used to model communication patterns between clusters, and the second was a graph of cliques that was used to model the interdependence of groups of clusters.

4.3.1 Directed Graph

In order to study the interactions between different clusters, a directed graph representing communications between clusters was formed as follows. The vertex set represented the 40 clusters identified using clustering techniques described in Section 4.2. We now extrapolate RQ1's definition of connections to clusters. For every pair of clusters c_i and c_j where an individual in cluster i tweeted an individual in cluster j, the graph has a corresponding directed edge ij. Tweets between individuals in the same cluster are represented by loops, that is, edges connecting a vertex (i.e. cluster) to itself in the directed graph.

Only tweets between clusters -- i.e. between agglomerations of users -- were included in the edges of the graph. Tweets in any of the three categories below were ignored:

1. @user messages (Type 1 in Section 4.1.2) addressed to users not found in any cluster (e.g. to a user whose metadata is not captured during data acquisition due to sampling limitations);

2. Retweets (Type 2 in Section 4.1.2) whose purpose is not for inter-tweeter discourse, but merely 'passing the message' akin to an email 'forward'; and

3. "Ether" tweets (Type 3 in Section 4.1.2) that are not directed to anyone in particular.

4.3.2 Results

There were 6353 tweets between clusters. The maximum number of tweets between any pair of clusters was 927 tweets from individuals in Cluster 0 to individuals in Cluster 0 (a loop in the graph). The maximum number of tweets between any pair of distinct clusters was 252 tweets from Cluster 11 to Cluster 0. The resulting directed graph had 423 unique edges.

The directed graph contained two components: the first component was a single isolated vertex corresponding to Cluster 38, and a second component containing all the other clusters. Individuals in Cluster 38 only communicate to individuals in their own cluster: there is no communication with (and hence influence on, or from) the outside world. Cluster 38 will be discussed in Section 5.1.

It would be expected that individuals in the same cluster would communicate with one another. This corresponds to a loop incident to the corresponding vertex in the directed graph. All vertices (and hence the cluster they represent) had loops except for the vertices corresponding to Clusters 22, 26, 27, 35, 36 and 39. In each of these clusters, there is no communication between members of the same cluster. A discussion of the reasons why no internal tweeting occurs between users in each of these cases is discussed in Section 5.2.

In directed graphs, it is interesting to note vertices which are *sources*, that is, have only outgoing edges (and possibly loops) and *sinks*, that is, vertices that only have in-going edges (and possibly loops). The isolated vertex corresponding to cluster 38 is trivially both a source and sink. The only edge incident to this vertex is a loop. There are no other source vertices in the graph. Hence, the underlying Twitter network has no identifiable non-trivial source clusters, that is, no group of users who only tweet out to other clusters while not receiving inward tweets.

There are two sink vertices in the graph. These vertices correspond to Clusters 27 and 35. Both these vertices have no loops. A discussion of the groups corresponding to these "sink" clusters is given in Section 5.3.

An analysis of the shortest distance between pairs of vertices in the graph is given in Table 1.

Each row in the table gives the minimum distance between the vertex indexed by that row and all other vertices. Rows with no green entries correlate with clusters that are closely connected to all clusters in the network. There are many vertices that are directly connected to each other via an edge. These are given a distance 1. In each of these cases the individual tweets directly to another individual. The longest shortest path is from Cluster 39 to Cluster 35: a directed path of length 29.

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39
0	0	1	2	1	2	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	2		2
1	1	0	4	2	10	3	5	4	7	2	9	2	7	1	3	15	6	4	13	8	16	18	6	11	4	12	3	16	18	14	18	17	5	19	1	19	16	1		16
2	2	1	0	3	13	4	6	5	8	3	12	10	16	9	11	23	7	12	18	15	17	21	20	14	12	19	4	24	24	22	24	25	1	25	26	25	1	4		24
3	1	2	2	0	3	1	2	1	1	2	2	2	2	1	1	2	1	1	1	1	2	1	1	1	1	3	3	1	1	1	1	1	1	2	2	3	1			3
4	2	3	12	5	0	6	8	7	10	3	17	19	15	4	11	23	9	12	18	16	21	25	14	1	12	20	6	24	24	24	24	22	13	25	25	25	24	6		24
5	1	2	3	1	9	0	2	1	1	2	8	2	6	3	2	5	1	3	1	7	1	1	5	10	3	1	2	6	3	1	1	1	4	4	2	2	6	2		6
6	2	3	9	5	15	6	0	1	7	3	14	2	12	4	8	5	1	9	3	13	6	8	11	16	9	17	6	6	8	4	8	7	10	9	8	9	6	6		6
7	1	2	2	1	2	1	1	0	1	2	1	1	1	1	1	1	1	2	2	2	1	1	1	2	3	2	1	2	2	1	3	1	1	1	2	2	2	2		2
8	1	2	2	1	3	1	2	1	0	2	2	2	2	1	1	2	1	2	1	2	1	1	1	1	4	1	1	3	3	1	3	1	3	4	2	4	3	2		3
9	2	3	11	5	17	1	7	6	9	0	16	3	14	4	10	6	8	11	2	15	7	9	13	18	11	1	6	7	4	10	9	8	12	5	9	11	7	6		7
10	1	2	2	4	1	5	2	1	2	2	0	7	5	3	1	5	1	2	1	6	3	2	4	1	2	2	5	6	6	1	6	4	3	7	2	7	1	5		6
11	1	2	2	1	3	1	2	1	2	2	2	0	2	1	1	3	1	2	1	1	4	2	1	4	2	1	2	4	1	2	1	1	3	1	2	2	1	2		4
12	1	2	2	1	4	2	2	1	4	2	3	2	0	3	1	2	3	2	3	1	1	2	1	5	2	4	2	3	3	3	1	1	3	4	4	4	3	2		3
13	1	1	2	1	8	1	2	1	1	2	7	1	5	0	1	2	1	2	2	6	1	2	4	9	2	1	2	3	1	1	1	3	2	1	2	3	2			3
14	1	2	1	1	2	1	1	1	1	2	1	1	2	1	0	1	1	1	1	1	1	1	1	1	2	1	2	1	2	1	2	3	2	1	2	2				2
15	1	2	4	1	3	1	3	2	2	2	2	2	7	3	3	0	1	4	1	1	1	2	6	4	4	1	2	1	1	1	1	5	1	2	1	1	2			1
16	1	2	2	1	2	1	2	1	1	2	1	1	2	1	1	1	0	2	1	3	2	1	3	2	1	2	1	2	2	2	1	3	3	2	3	2	2			2
17	2	3	7	5	13	6	8	7	10	3	12	2	4	4	1	9	9	0	11	1	5	11	3	14	11	6	6	10	1	10	13	12	8	2	13	2	8	6		10
18	1	2	2	1	2	1	2	1	1	2	1	1	5	3	1	2	1	2	0	6	2	2	4	3	2	1	2	3	3	1	1	1	3	4	1	4	1	2		3
19	1	2	2	1	2	1	2	1	1	2	1	1	1	3	1	3	3	2	2	0	1	1	4	1	2	1	2	4	4	2	4	1	3	5	2	1	4	2		4
20	1	2	2	1	3	1	2	1	4	2	2	2	2	1	1	5	3	2	1	1	0	7	1	4	2	1	2	6	3	6	2	1	3	4	2	7	6	2		6
21	1	2	2	1	2	2	2	1	2	2	1	6	2	3	1	9	1	2	5	1	10	0	1	3	2	4	2	10	7	1	2	1	3	8	2	2	10	2		10
22	1	2	2	4	4	1	2	1	2	2	3	9	1	3	1	7	1	2	8	2	3	1	0	5	2	4	5	8	10	6	10	2	3	11	3	11	10	5		8
23	1	2	6	1	3	2	2	1	4	2	2	4	9	3	6	3	1	2	2	8	6	3	1	0	6	1	2	3	3	1	3	1	7	4	4	4	3	2		3
24	1	2	2	1	8	2	4	3	6	2	7	12	5	3	1	15	5	2	11	6	16	18	4	9	0	10	2	16	13	19	18	17	3	14	18	20	16	2		16
25	1	2	1	1	3	1	2	1	1	2	2	2	1	1	1	2	1	2	1	1	1	2	1	4	2	0	2	3	3	1	1	1	2	4	1	4	3	2		3
26	1	2	5	4	11	1	1	2	3	2	10	15	8	3	4	18	2	5	14	9	19	21	7	12	5	13	0	19	16	22	21	20	6	17	21	23	19	5		19
27																												0												
28	1	2	9	2	15	3	5	4	7	2	14	19	12	1	8	23	6	9	18	13	24	21	11	16	9	17	3	24	0	22	20	25	10	1	26	1	24	3		24
29	1	2	1	1	2	1	2	1	1	2	1	2	2	3	1	1	3	2	1	1	2	1	1	2	2	2	2	3	2	0	2	1	2	1	2	1	2	2		2
30	1	2	2	1	4	2	4	3	6	2	3	3	1	3	1	6	5	2	2	2	7	1	4	5	2	1	2	7	4	1	0	1	3	5	2	2	7	2		7
31	1	1	2	1	2	1	1	1	1	2	1	2	1	1	1	1	2	1	1	1	1	1	2	1	2	2	1	1	1	1	1	0	3	2	1	2	2	2		2
32	1	2	11	4	14	5	7	6	9	2	13	17	2	3	10	4	8	11	18	16	3	2	1	15	11	1	5	5	5	3	5	6	0	6	7	6	12	5		5
33	1	2	2	4	8	1	3	2	5	2	7	3	5	3	1	1	4	2	2	6	2	5	4	9	2	1	5	2	4	6	4	3	3	0	4	5	2	5		2
34	1	2	6	4	2	5	2	1	4	2	1	2	9	3	5	5	3	6	1	10	6	2	8	3	6	1	5	6	3	3	2	1	7	4	0	4	6	5		6
35																																				0				
36	1	2	1	4	2	5	2	1	4	2	1	1	4	3	5	1	3	6	5	5	2	4	3	3	6	4	5	2	2	5	4	3	2	3	1	6	0	5		2
37	1	2	11	4	17	5	7	6	9	2	16	2	14	3	10	5	8	11	1	15	6	8	13	18	11	19	5	6	3	9	8	7	12	4	8	10	6	0		6
38																																							0	
39	1	2	11	4	17	5	7	6	9	2	16	21	14	3	10	24	8	11	20	15	25	27	13	18	11	19	5	25	22	28	27	26	12	23	27	29	25	5		0

Table 1. Shortest distances between vertices (representing clusters) in directed graph. When no directed path connects two vertices, the cell is left blank.

4.3.3 Graph of Cliques

A *clique* is a set of vertices where every pair of vertices in the set is connected by a bi-directed edge. We created a second graph which we refer to as a *graph of cliques*, by identifying the maximal cliques in the directed graph and representing each maximal clique by a vertex in this graph. Two vertices are connected by an edge in the graph of cliques if they contain some clusters in common. The graph of cliques is undirected. We ignored loops as every vertex trivially contains clusters in common with itself.

4.3.4 Results

The graph of cliques had 44 vertices representing the 44 maximal cliques (excluding the 16 maximal cliques that were trivial, that is, a single vertex). The maximum possible number of edges in an undirected graph with 44 vertices that has no loops is $\binom{44}{2} = 946$. In the graph of cliques formed from this Twitter network, all but 167 of these edges were present - almost 83% of all possible edges.

Seven of the cliques were connected by edges to all other cliques. The sets of clusters belonging to each of the seven cliques is given in Table 2. Cluster 14 is common to all of these sets of clusters. Clusters 21 and 31 are each in all but two of the cliques in Table 2 (albeit in different sets). We discuss characteristics of these clusters in Section 5.4.

| {3, 7, **14**, 19, 20, 25, 31} |
| {3, 7, 13, **14**, 20, 25, 31} |
| {7, **14**, 19, 21, 31} |
| {**14**, 16, 21, 22} |
| {**14**, 16, 21, 31} |
| {**14**, 17} |
| {**14**, 21, 29, 31} |

Table 2. Cliques (sets of clusters) adjacent to all other cliques in Graph of Cliques

4.3.5 Addressing RQ2

RQ2. Using graph-theoretic approaches, can we identify emergent communication patterns between clusters of Twitter users?

In this section, we identified groups of Twitter users that were isolated from the rest of the network. We also identified groups that did not communicate outside their cluster but did communicate within their cluster (corresponding to sink vertices in the network).

An analysis of the directed graph summarised in Table 1 gives some idea of the minimum length of tweets that would be required for one cluster to communicate with another using existing communication pathways.

The graph of cliques enabled us to identify a cluster of users that are central to our Twitter network.

In Section 5, we examine the clusters identified and the characteristics and properties of these users in order to explain the nuances in communication patterns of these groups.

5 Analysis of Interesting Clusters

In order to profile the users related to clusters that we identified as interesting, we used the original metadata in our main corpus. For example, some tweets can be clearly identified as coming from a media source, some from a politician, and some appear to be from a member of the public.

Using the Tableau BI suite, we filter out certain clusters of interest and observe the most-frequent tweeters to those clusters, based on volume of tweets found in the corpus.

We utilise metadata attached to these users to infer who they are. This can be achieved by studying their self-bio description, their full name and nickname, and Twitter username, as gathered from our dataset. If more information is needed (e.g. due to the absence of meaningful metadata, or the vagueness of provided information), we explore that particular user's Twitter feed (including profile, posts, images, and links) on the Twitter website itself.

In addition, if they are public figures, we can search their Twitter profile and mentions of their username via the Twitter search facility and Google to look up more information about that public figure.

5.1   Isolated Clusters

The isolated cluster, Cluster 38 (Figure 5) contains users who, based on their self-description, self-identify as active contributors to the community based on their occupation. From the most active users (by volume of tweets) in Cluster 38, we find that they are in the health, legal, and educational professions.
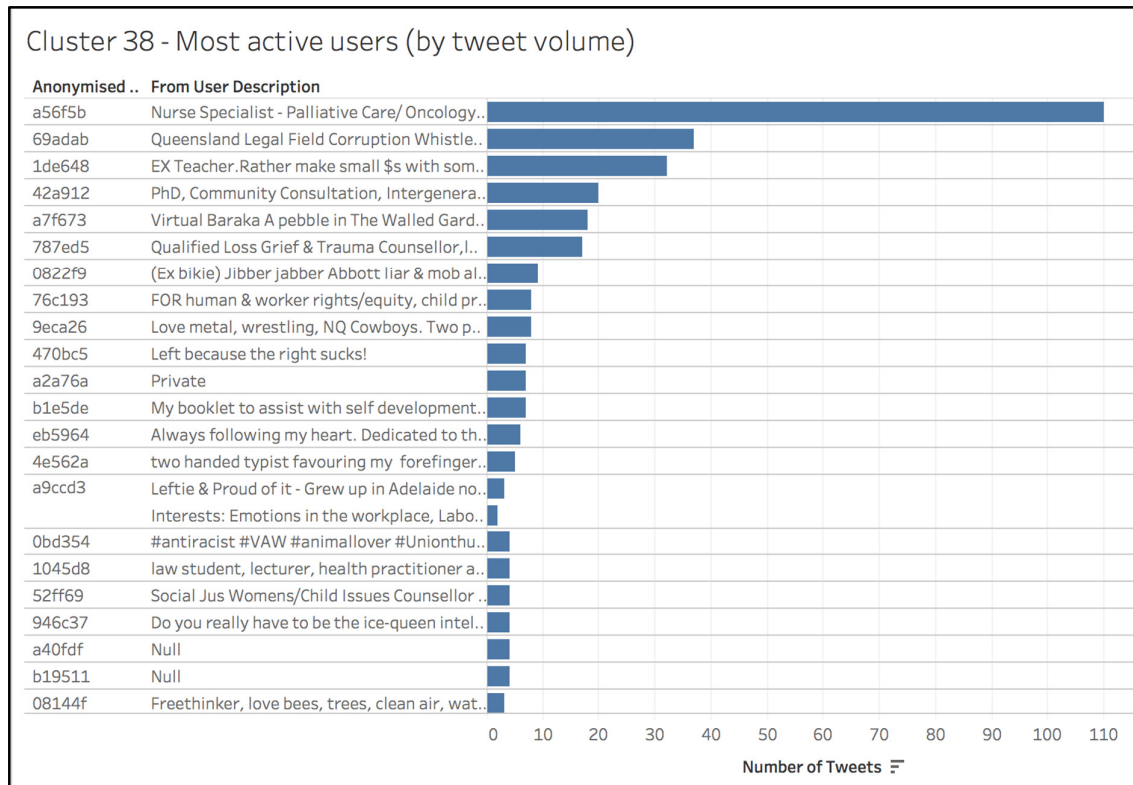

Figure 5. Tableau visualisation of the most active users in Cluster 38 (the isolated cluster), including their User Description on Twitter as well as tweet volume. Names of the users are coded to ensure anonymity.

These active users exhibit traits of liberalism or progressiveness based on their self-description. A recent market research survey on Australian Twitter users indicate that Australian Twitter users tend to be progressive or liberal, inferred based on their political leanings (Roy Morgan Research, 2016). Our results suggest that this cluster has no influence on other clusters in our network.

5.2   Clusters with no Internal Communication

The vertices representing Clusters 22, 26, 27, 35, 36 and 39 had no incident loops, that is, there were no tweets from each individual cluster to itself. Users in these clusters are tweeting only to, and receiving tweets from, users in other clusters. The most frequent tweets in these clusters were from the NSW Teachers Federation (see Figure 6). Other active users in these clusters are also education-focused.
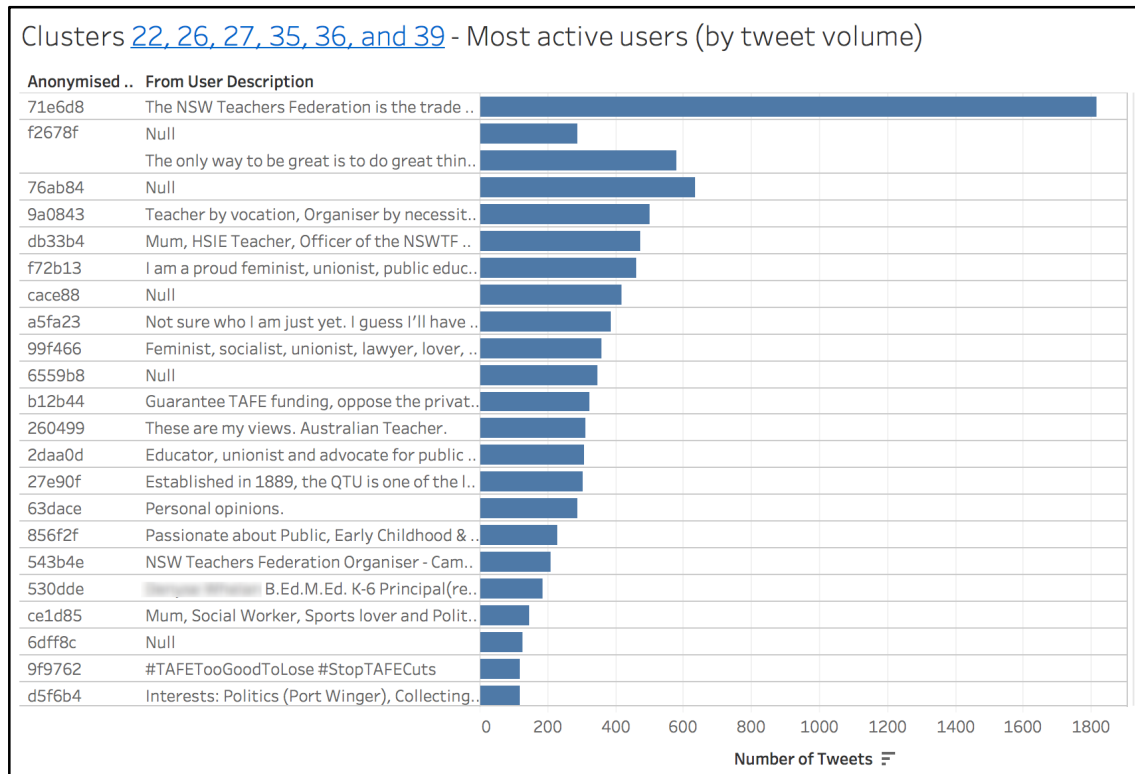
Clusters 22, 26, 27, 35, 36, and 39 - Most active users (by tweet volume)

Anonymised ..	From User Description	Number of Tweets
71e6d8	The NSW Teachers Federation is the trade ..	
f2678f	Null	
	The only way to be great is to do great thin..	
76ab84	Null	
9a0843	Teacher by vocation, Organiser by necessit..	
db33b4	Mum, HSIE Teacher, Officer of the NSWTF ..	
f72b13	I am a proud feminist, unionist, public educ..	
cace88	Null	
a5fa23	Not sure who I am just yet. I guess I'll have ..	
99f466	Feminist, socialist, unionist, lawyer, lover, ..	
6559b8	Null	
b12b44	Guarantee TAFE funding, oppose the privat..	
260499	These are my views. Australian Teacher.	
2daa0d	Educator, unionist and advocate for public ..	
27e90f	Established in 1889, the QTU is one of the l..	
63dace	Personal opinions.	
856f2f	Passionate about Public, Early Childhood & ..	
543b4e	NSW Teachers Federation Organiser - Cam..	
530dde	B.Ed.M.Ed. K-6 Principal(re..	
ce1d85	Mum, Social Worker, Sports lover and Polit..	
6dff8c	Null	
9f9762	#TAFETooGoodToLose #StopTAFECuts	
d5f6b4	Interests: Politics (Port Winger), Collecting..	

Figure 6. Tableau visualisation of the most active users in Clusters 22, 26, 27, 35, 36, and 39 (clusters with no internal communication), including their User Description on Twitter as well as tweet volume. Names of the users are coded to ensure anonymity. Other personally-identifiable information in descriptions is redacted.

This possibly indicates that this cluster identifies group(s) who are using Twitter as a forum to discuss education rhetoric with a broader community. From these 'no-loop' clusters, we agglomerated the users found within these clusters and studied the descriptions of the top users in the agglomeration. The topmost users (by frequency) self-describe their vocations as teachers and educators. By analysing their description, these 'no-loop' cluster users tend to be progressive and self-describe as feminists and proud union members. However, based on our graph cluster analysis, their communication patterns are different compared to Cluster 38 (Section 5.1) -- Cluster 38's communications are all internal, but this group of clusters are only external. In other words, Cluster 38 keeps information for users within the cluster, while Clusters 22, 26, 27, 35, 36 and 39 are for dispersing information to users outside their current cluster. These clusters with no internal communication can be said to exhibit 'call-to-action' behaviour – examples would be tweets calling for support for a particular party's initiative or to promote awareness of a particular policy.

5.3 Sink Clusters

Vertices representing Clusters 27 and 35 are sinks with no loops in the directed graph. A loopless sink vertex in the directed graph represents a cluster where all tweets are coming to users in the clusters, that is, there are no outgoing tweets and no tweets between individuals in the cluster. Thus, these clusters identify users that are receivers only. Clusters 27 and 35 are a subset of the clusters discussed in Section 5.2.

In Figure 7, we see the demographic breakdown is similar to Figure 6 even when we restricted to these clusters. By agglomerating Clusters 27 and 35 (being a subset of the 'no-loop' clusters), we also notice similar characteristics to the superset in Section 5.2, where users are mainly teachers who are progressive. However, the differences between the frequency distributions in Figures 6 and 7 exist only after the top 20th user (i.e. the user coded as "6dff8c" in Figure 7).
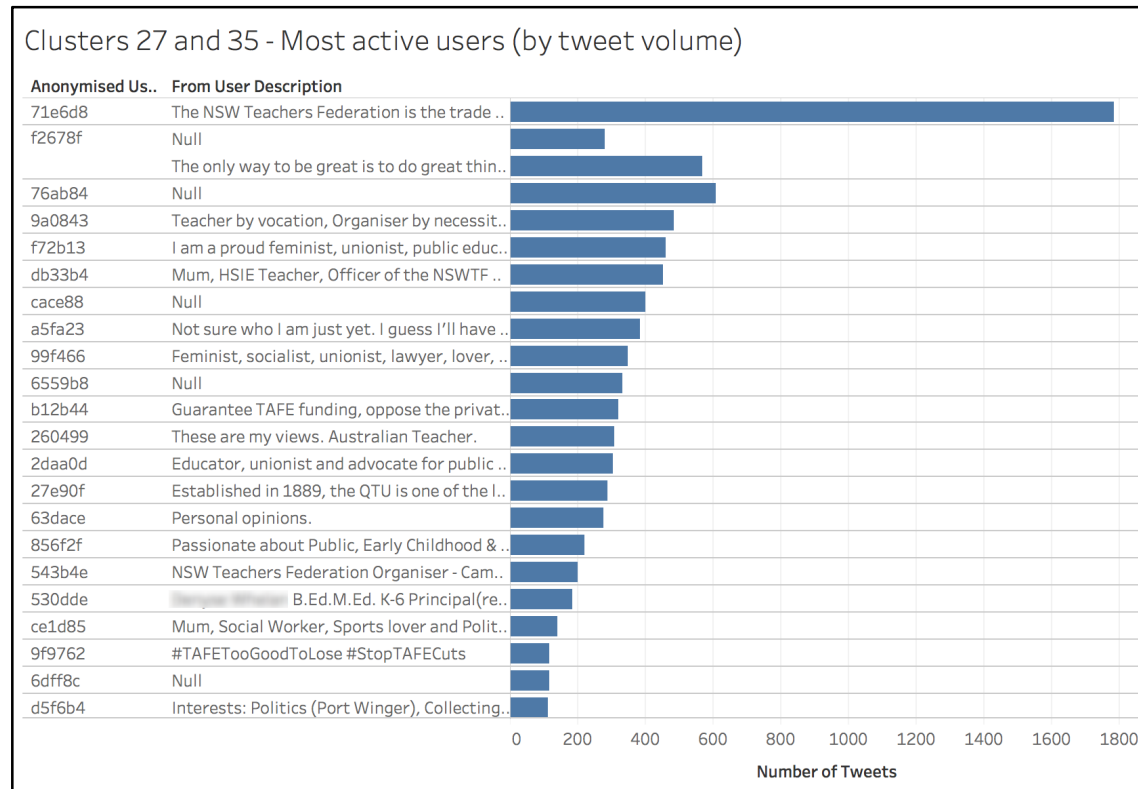

Figure 7 data — Clusters 27 and 35 – Most active users (by tweet volume):

Anonymised Us..	From User Description	Number of Tweets (approx.)
71e6d8	The NSW Teachers Federation is the trade ..	~1800
f2678f	Null	~290
	The only way to be great is to do great thin..	~560
76ab84	Null	~610
9a0843	Teacher by vocation, Organiser by necessit..	~480
f72b13	I am a proud feminist, unionist, public educ..	~460
db33b4	Mum, HSIE Teacher, Officer of the NSWTF ..	~450
cace88	Null	~400
a5fa23	Not sure who I am just yet. I guess I'll have ..	~390
99f466	Feminist, socialist, unionist, lawyer, lover, ..	~350
6559b8	Null	~340
b12b44	Guarantee TAFE funding, oppose the privat..	~320
260499	These are my views. Australian Teacher.	~310
2daa0d	Educator, unionist and advocate for public ..	~310
27e90f	Established in 1889, the QTU is one of the l..	~300
63dace	Personal opinions.	~280
856f2f	Passionate about Public, Early Childhood & ..	~230
543b4e	NSW Teachers Federation Organiser - Cam..	~210
530dde	B.Ed.M.Ed. K-6 Principal(re..	~190
ce1d85	Mum, Social Worker, Sports lover and Polit..	~150
9f9762	#TAFETooGoodToLose #StopTAFECuts	~120
6dff8c	Null	~110
d5f6b4	Interests: Politics (Port Winger), Collecting..	~100

Figure 7. Tableau visualisation of the most active users in Clusters 27 and 35, including their User Description on Twitter as well as tweet volume. Names of the users are de-identified to ensure anonymity. Other personally-identifiable information in descriptions is redacted.

5.4 Important Clusters in the Graph of Cliques

The graph of cliques contained seven maximal cliques that were connected by edges to all other maximal cliques in the graph (see Section 4.3.4). Cluster 14 was in each of these seven cliques. This universal cluster could represent the most engaged users in our network. Cluster 14 contains 288 users, with each user only contributing 6 tweets or less, which is far less than those found in the previous clusters discussed. It is difficult to conclusively describe features common to most of the top users in this cluster, as the descriptions contain an eclectic mix of terms and descriptors, Tableau highlighted an interesting anomaly - the top users (seen in Figure 8) frequently change/edit their self-written descriptions (bio) on Twitter, as illustrated by users with more than one "From User Description" metadata item. It appears that our methods have identified an engaged, freethinking and genuine user group.
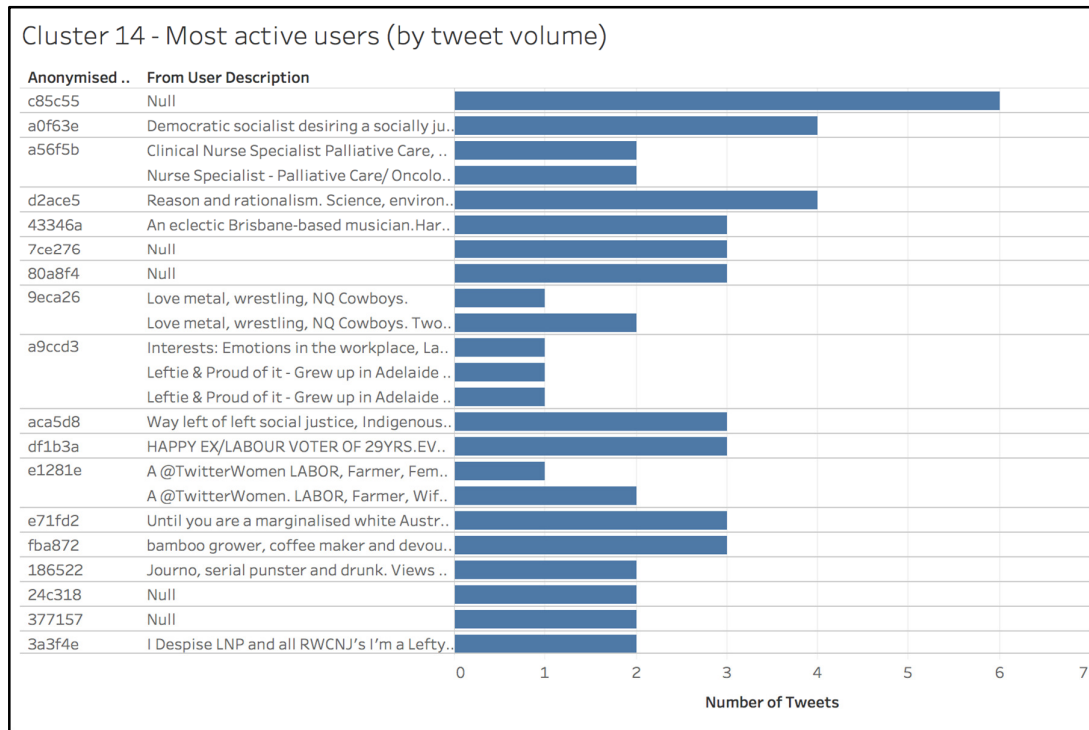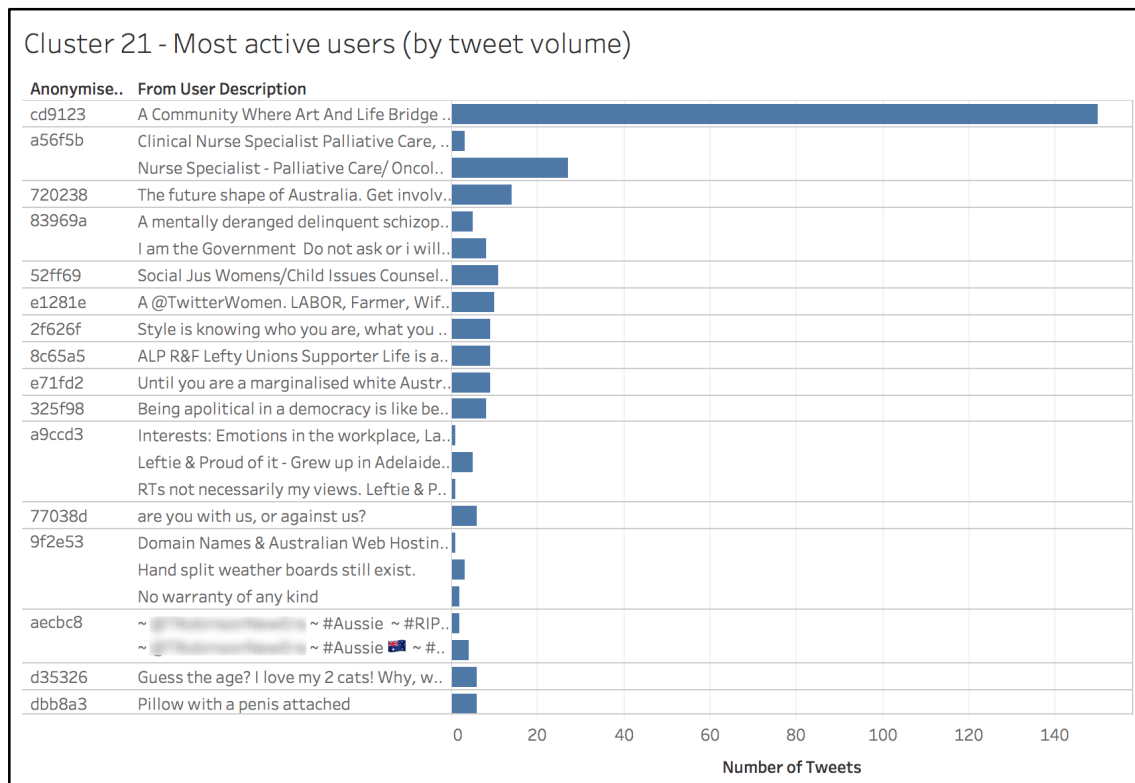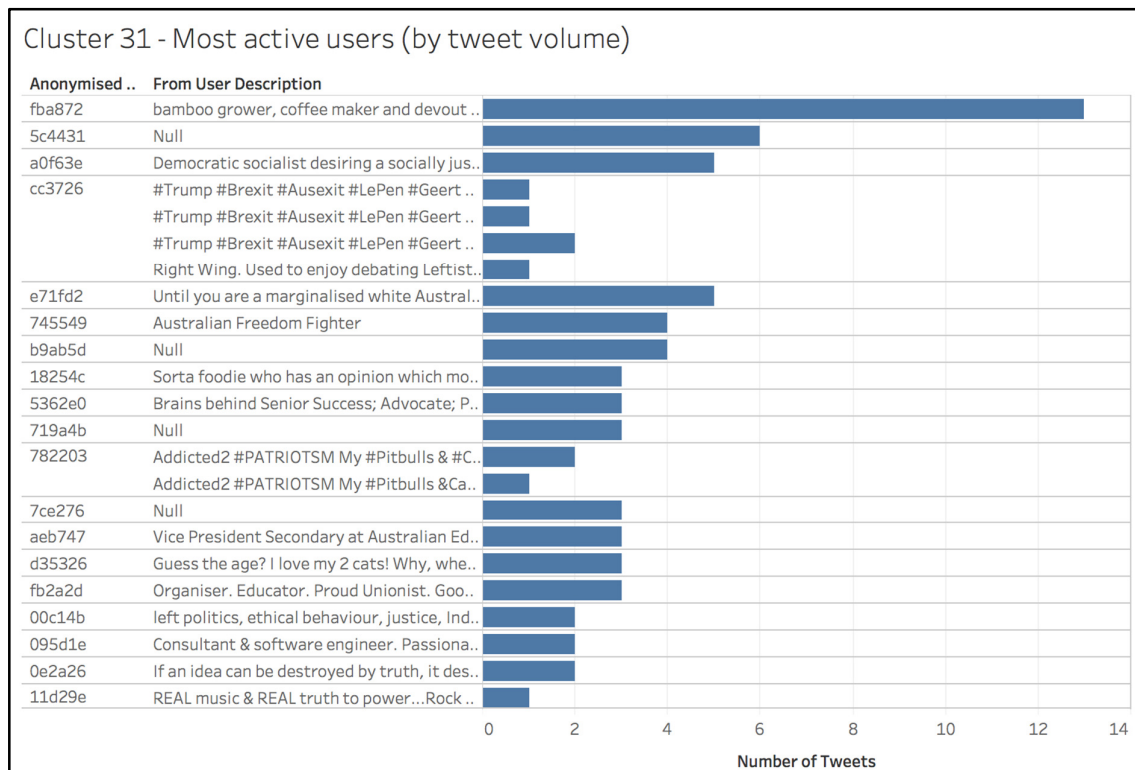
Figure 8. Tableau visualisation of the most active users in Cluster 14, including their User Description (bio) on Twitter as well as tweet volume. Names of the users are coded to ensure anonymity.

Cluster 21 and 31 were also identified as interesting in Section 4.3.4. Cluster 21 exhibits similar traits to Cluster 14, such as an eclectic mix of user descriptions and frequent editing of their descriptions on Twitter. However, an outlier of interest in Cluster 21 is the user coded as "cd9123", who tweeted a disproportionate number of tweets (150) - nearly 5 times as many as the second most active tweeter in this cluster. All other tweeters tweeted a maximum of 20 times.

Figure 9. Tableau visualisation of the most active users in Cluster 21, including their User Description on Twitter as well as tweet volume. Names of the users are de-identified to ensure anonymity. Other personally-identifiable information in descriptions are redacted.

Finally, we analyse the trends of users in Cluster 31. Again, similar to individual analyses of Clusters 14 and 21, an eclectic mix of user descriptions and frequent changing of user description on Twitter is found in the top user list. However, the distribution of tweets is similar to Cluster 14, as the maximum number of tweets per user detected in this cluster is 13, while all others have 6 or less tweets.

Figure 10. Tableau visualisation of the most active users in Cluster 31, including their User Description on Twitter as well as tweet volume. Names of the users are de-identified to ensure anonymity.

Clusters 21 and 31 are in cliques that are almost as connected as those containing Cluster 14. The similar behaviour of the nodes in the Graph of Cliques (see Section 4.3.4) is mirrored in the eclectic behaviour of the users within the corresponding clusters.

6   Conclusions and Further Work

We have demonstrated that a methodology using a combination of clustering and graph theoretical techniques can provide significant insight into Twitter user interactions in relation to topical issues. This methodology can be applied to any specific topics or combination of topics trending on social media. Our analysis of cliques based on clustered features of tweet content, by definition, is unique compared to social network research utilising user ties alone (Quintaine & Kleinbaum, 2011; Kwak, Lee, Park & Moon, 2010).

Using this methodology, we have established that, in the context of our particular case study, cohorts of the general public that we could identify as being the 'common man', have very limited presence in the wider discourse. The formation of cliques indicate that the identified cohorts exhibit strong *homophily* – in other words, such "…networks are homogeneous with regard to many sociodemographic, behavioural, and intrapersonal characteristics" (McPherson, 2001). Homophily in this regard is also found in other online network settings (Quintaine & Kleinbaum, 2011). In this study, the strong homophily in certain groups (e.g. political and environmental) appears to have had the greatest influence over the network structure. We conclude that the 'common man' does not have a voice in this particular network.

There are some exceptions to this. For example, users identified as teachers mainly belonged to one of two clusters. The tweeters in one cluster mainly communicated to users within their own cluster regarding interest internal to their group. Conversely, the tweeters in the second cluster predominantly tweeted to external users regarding topics of more widespread interest.

6.1 Further work.

Our case study has only considered a single Twitter network observed over a window in time. Applying this approach to multiple contexts including other media and/or other issues of discourse would provide further information. Possibly the 'common man' has a voice when the domain of discourse is not of interest to large cohorts of users. Another direction of future research is to investigate the long-term influence of groups of users in the Hazelwood discourse.

An alternative approach to the analysis conducted in this research, would be to cluster based on the descriptions of the users in the sample. However, the descriptions may not be accurate or consistent as many users prefer anonymity.

Acknowledgements

References

Abbas, O.A. (2008) Comparisons between data clustering algorithms. *The International Arab Journal of Information Technology*, 5(3), 320-325.

Aggarwal, C. C., & Zhai, C. (2012). A Survey of Text Clustering Algorithms. In *Mining Text Data* (pp. 77–128).

Aittokallio, T., & Schwikowski, B. (2006). Graph-based methods for analysing networks in cell biology. *Briefings in Bioinformatics*, 7(3), 243–255.

Anderson, S. (2017, March 30). Everything you need to know about the Hazelwood shutdown. Retrieved September 1, 2017, from http://www.abc.net.au/news/2017-03-30/hazelwood-power-plant-shutdown-explained/8379756

Bang-Jensen, J., & Gutin, G. Z. (2008). *Digraphs: Theory, Algorithms and Applications*. Springer.

Barabási, A.-L. and R. Albert, Science **286**, 509 (1999).

Bar-Yam, Y. (2011). Concepts: Power Law | NECSI. Retrieved October 4, 2017, from http://www.necsi.edu/guide/concepts/powerlaw.html

Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An Open Source Software for Exploring and Manipulating Networks. In *International AAAI Conference on Weblogs and Social Media*. Retrieved from http://www.aaai.org/ocs/index.php/ICWSM/09/paper/view/154

Batagelj, V., & Mrvar, A. (2004). Pajek — Analysis and Visualization of Large Networks. In *Mathematics and Visualization* (pp. 77–103).

Bearman, P. S., Moody, J., & Stovel, K. (2004). Chains of Affection: The Structure of Adolescent Romantic and Sexual Networks. *The American Journal of Sociology*, *110*(1), 44–91.

Bernard, H., Killworth, P. D. & Sailer, L. (1980). Informant accuracy in social network data IV: a comparison of clique-level structure in behavioral and cognitive network data. *Social Networks*, Elsevier BV, 2, 191-218.

Bilgili, M., Ozbek, A., Sahin, B., & Kahraman, A. (2015). An overview of renewable electric power capacity and progress in new technologies in the world. *Renewable and Sustainable Energy Reviews*, *49*, 323–334.

Borra, E., & Rieder, B. (2014). Programmed method: developing a toolset for capturing and analyzing tweets. *Aslib Journal of Information Management*, *66*(3), 262–278.

Bunn, A. G., Urban, D. L., & Keitt, T. H. (2000). Landscape connectivity: A conservation application of graph theory. *Journal of Environmental Management*, *59*(4), 265–278.

Byrnes, L., Brown, C., Foster, J., & Wagner, L. D. (2013). Australian renewable energy policy: Barriers and challenges. *Renewable Energy*, *60*, 711–721.

Cheong, M. (2013). *Inferring social behavior and interaction on twitter by combining metadata about users & messages* (PhD). https://doi.org/10.4225/03/58b5009d3726a

Cormode, G., Krishnamurthy, B., & Willinger, W. (2010). A manifesto for modeling and measurement in social media. *First Monday*, *15*(9). https://doi.org/10.5210/fm.v15i9.3072

Davies, P. A. (2014). Green crime and victimization: Tensions between social and environmental justice. *Theoretical Criminology*, *18*(3), 300–316.

Diestel, R. (2005). *Graph Theory*. Springer Science & Business Media.

Dudley-Nicholson, J. (2016, May 27). Twitter to help decide Australian federal election with most uncommitted. Retrieved September 27, 2017, from http://www.news.com.au/technology/twitter-may-help-decide-who-rules-australia-with-millions-of-votes-up-for-grabs-in-the-online-election/news-story/d2304b039651b86fe203f8dab83837bb

Evans, G., & Phelan, L. (2016). Transition to a post-carbon society: Linking environmental justice and just transition discourses. *Energy Policy*, *99*, 329–339.

Freeman, L. C. (1977). A Set of Measures of Centrality Based on Betweenness. *Sociometry*, *40*(1), 35.

Furlan, C., & Mortarino, C. (2017). Forecasting the impact of renewable energies in competition with non-renewable sources. *Renewable and Sustainable Energy Reviews*. https://doi.org/10.1016/j.rser.2017.05.284

Green, D. G. (2014). *Of Ants and Men*. Berlin, Heidelberg: Springer Berlin Heidelberg.

Holmgren, A. J. (2006). Using graph models to analyze the vulnerability of electric power networks. *Risk Analysis: An Official Publication of the Society for Risk Analysis*, *26*(4), 955–969.

Hull, C. (2017, April 27). Turnbull's foolish gamble to pander to Big Fossil over Adani coal mine. Retrieved September 1, 2017, from https://www.brisbanetimes.com.au/opinion/malcolm-turnbulls-foolish-gamble-to-pander-to-big-fossil-over-adani-coal-mine-20170427-gvtusc.html

Jansen, B.J., Sobel, K and Cook, G. (2011). Classifying ecommerce information sharing behaviour by youths on social networking sites. *Journal of Information Sciences, 37*, 120-136.

Johnstone, P., & Hielscher, S. (2017). Phasing out coal, sustaining coal communities? Living with technological decline in sustainability pathways. *The Extractive Industries and Society*. https://doi.org/10.1016/j.exis.2017.06.002

Knott, M. (2017, May 12). "We have to face reality": Greens prepared to strike school funding deal with the Coalition. Retrieved September 1, 2017, from http://www.smh.com.au/federal-politics/political-news/we-have-to-face-reality-greens-prepared-to-strike-school-funding-deal-with-the-coalition-20170512-gw3m31.html

Kolokotsa, D. (2017). Smart cooling systems for the urban environment. Using renewable technologies to face the urban climate change. *Solar Energy*. https://doi.org/10.1016/j.solener.2016.12.004

Kwak, H., Lee, C., Park, H., & Moon, S. (2010). What is Twitter, a social network or a news media? *Proceedings of the 19th International Conference on World Wide Web (WWW '10)*. http://dx.doi.org/10.1145/1772690.1772751

Mason, O., & Verwoerd, M. (2007). Graph theory and networks in Biology. *IET Systems Biology*, *1*(2), 89–119.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology, Annual Reviews*, 27, 415-444.

Moreno, J. L. (1934). *Who shall survive?: A new approach to the problem of human interrelations*. Nervous and mental disease publishing co.

Puddy, R. (2015, December 21). South Australia faces crisis over power availability, prices. Retrieved September 1, 2017, from http://www.theaustralian.com.au/business/mining-energy/south-australia-faces-crisis-over-power-availability-prices/news-story/16855143aae5e9eebcccfee2ea043673

Quintane, E. & Kleinbaum., A. (2011). Matter Over Mind? E-mail Data and the Measurement of Social Networks. *Connections*, 31, 22-46.

Roy Morgan Research. (2016, June 21). Twitter: a bird in the hand for 5.4m Australians. Retrieved October 5, 2017, from http://www.roymorgan.com/findings/6859-twitter-active-users-australia-march-2016-201606210913

Sen, R., & Bhattacharyya, S. C. (2014). Off-grid electricity generation with renewable energy technologies in India: An application of HOMER. *Renewable Energy*, *62*, 388–398.

Shamma, D. A., Kennedy, L., & Churchill, E. F. (2009). Tweet the debates. *Proceedings of the first SIGMM workshop on Social media - WSM '09*. https://doi.org/10.1145/1631144.1631148

Twidell, J., & Weir, T. (2015). *Renewable Energy Resources*. Routledge.

Wasserman, S., & Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge University Press.

Xu, R., & Wunsch, D., 2nd. (2005). Survey of clustering algorithms. *IEEE Transactions on Neural Networks / a Publication of the IEEE Neural Networks Council*, *16*(3), 645–678.

Zhang, S., Wang, R., and Zhang, X. Identification of overlapping community structure in complex networks using fuzzy c-means clustering. (2007). *Physica A: Statistical Mechanics and its Applications,* 374(1): 483-490.