When Trump tweets does the world catch a cold?

A detailed investigation into the online activity of Donald Trump via the social media platform Twitter, and whether it is causing investor behaviour to change, and markets to fluctuate all the way to Australia.

- **Research and report produced for:** Assignment 2, COSC2671 (Social Media and Network Analytics)
- **Research and report conducted by**:
- **Date submitted:**

Table of Contents

Introduction:

US President Donald J. Trump is a prolific user of Twitter, averaging close to 10 tweets per day since taking office in 2017. Ranging from the incoherently absurd to downright warmongering, The Donald's tweets come thick, fast and without apology.

This report endeavours to present the research and analysis we have conducted into the relationship between Trump's tweets and the performance of the Australian stock market between April and September 2019.

Our overarching research question was *'is there conclusive evidence that Trump's tweets are felt across the globe, do reach Australia, and do affect investor confidence and activity, causing fluctuations in what would otherwise be relatively stable ASX listed share prices?'*.

From our own firsthand experiences, we understand the purpose and design behind a social media platform – an online platform on which people, friends, foes or otherwise, can engage and interact with each other, directly or indirectly.
The cogs and wheels of social networks are also instinctive enough to us, as they are as natural to humans as walking and talking. The ability however of being able to detect and measure the latter from the former is quite a bit more abstract.

In theory perhaps the idea makes sense, one could argue why wouldn't the online versions of ourselves cluster around ideologies, form communities and influence one another in the same ways we do in reality. In practice, it's evidently not that simple, yet this is the task we have set ourselves, to use social media interactions and actions of an individual to detect, measure and draw conclusions about events occurring in the real world, very real events causing a great number of people significant concern.

The approach we have used to investigate our research question and gain insights into this relationship was to break it into two main parts: Part A and part B.

Part A – The Interaction:

This part was designed to explore the virtual social network of Donald Trump, his online community, his followers, and ideally who he influences with his tweets. Part A aimed to understand the pathways on which the content Trump produces flows, and to whom it flows to and through. From this we hypothesised we would be able to draw conclusions on how these tweets were able to enter the decision-making cognition of investors here on Australian soil.

Part B – The Action:

This second stage focused on examining the content produced by Trump, what we referred to as the 'action'. We hypothesised that the content, frequency, and timing of his tweets would provide conclusive insight into which type of Trump twitter action could and would cause financial market turbulence. This phase of our research involved the scrapping and analysis of 3,000 tweets from Trump over a period of 6 months. Parallel analyses was conducted on tweets produced by Commsec to help us draw conclusions that Trumps

tweets were indeed making their way through the decision making turnstiles of Australia's key sharemarket players.

Background:

The idea that a Trump tweet can move markets isn't new.

Research into DT's tweets for to gain insights and conclusive evidence of market impact is not our brainchild. Multiple large institutions such as Citibank, Bank of America, and JP Morgan have led the way in identifying the correlation between his tweets and fluctuations in various financial markets. Their hope is that their research, aided by machine learning can begin to establish causation patterns, predict these fluctuations and give them an edge on the investment stage.

As reported by The Age [1] one such example is the **Volfefe index**, a project developed by JP Morgan "which seeks to [classify and] quantify the impact of Trump's 10,000-plus tweets since winning office" on volatility in interest rate markets. The report goes on suggesting that "as with other research by Citibank and Bank of America, JP Morgan found that both the volume of tweets and some predictable keywords impact prices and volatility."

The Report goes on to state that "it was predictable – and observable -- that volatility would spike when Trump announces new tariffs on China's exports or increased rates on existing tariffs. It's more difficult, however, to anticipate the magnitude of those effects."

"Citigroup has been looking at the effect of the tweets on currency markets and Bank of America has been analysing their equity market effects and come to similar conclusions.

The simple conclusion from the work of the three investment banks is that Trump's tweets have increased uncertainty, volatility and risk in US bond, equity and currency markets.

Our research aims to expand on these initial ideas, investigating the affects beyond the US borders, into Australia.

Part A: The interactions

Purpose:

The purpose of Part A was to explore and gauge the importance of how, through whom, and to whom Trumps tweets flow, i.e. which pathways and communities, if any, are providing slipstreams and megaphones for his diatribe to flow all the way to Australia, messing with our daily trading activities.

The purpose of Part A is to provide a foundation of 'who he can influence', from which we can then begin to explore 'how and with what he influences them' further in Part B.

With this information we believed we could conclude, or at least infer, defined pathways, and obvious communities and other key players which we could deduce as acting as pseudo-verifiers, that give the tweets apparent credibility.

Logic

As market fluctuations caused by Trump is not a new phenomenon, we expected to see a significant representation of key financial players from across the globe, following Trump, retweeting Trump, and at the very least form a noticeable community connected to Trump.

1. Get an overview of Trump's Twitter profile/account
2. Explore the followers of DT
3. Explore who DT followers, defined as 'friends'
4. Extend this list into the followers of the 47 friends trump follows.
5. This investigation was done to observe the accounts that notice his actions on Twitter, evaluate whether their owners are key people in the Australian or American financial markets, and analyse if their Twitter activity may be influenced by Trump's activity.

Data Retrieval:

We determined that the initial data needed for the interactions analysis needed to include as many entities and records as possible from the full list of Donald Trump's followers, friends, and tweets, as well as key players from financial and business sectors. This determination was based on the assumption that Trumps 1st degree community (his direct followers, and those that he follows), as well as the list of key financial players would include well known and influential people in their own right.

Trumps Followers & Friends:

The Twitter data was retrieved using the Twitter API and the Tweepy python package, while the list of key financial players was constructed manually, from which approximately 150 twitter accounts of key international businesses, financial institutions, and persons were sourced. The list was based on a recent report by Business Insider [2].

Our first major challenge came in retrieving the 65 million followers of Donald Trump. It was not possible due to the limitations of the Twitter API - a sample of followers was instead obtained. The account's followers list was queried until the rate limit was reached, resulting in a sample size of approximately 10,000 followers – only 0.015% of the total number of followers, but still a useful sample to analyse.

Unlike the number of followers of Donald Trump, the number of friends (i.e. accounts who Trump follows) is only 47, and no rate limit was exceeded in retrieving these.

The list of 10,000 followers, and 47 friends were obtained to evaluate Trump's first-degree Twitter connections.

Trumps Tweets:

To obtain Trump's tweets, a similar approach was taken, querying Donald Trump's account for tweets until the rate limit was reached, yielding 3,000 tweets dated between the April 22 to September 27, 2019 – unfortunately again limited by the Twitter access, but nonetheless a useful sample for our analysis.

Regarding Part A, the purpose of retrieving the 3,000 tweets was to use them to extract the account IDs Trump mentions in them, along with the IDs of the authors of the tweets he replies to and retweets.

To evaluate who Trump directly interacts with on Twitter, the user mentions in his tweets and the users he replies to in his tweets were extracted from the existing sample of 3,000 tweets, ensuring the accounts were identified by the Twitter screen name. This was done to establish whether he has a closer relationship, measured by frequency of interactions, than to any of his friends and followers. If that was to be the case, and if any of the close relationships established were with key financial people, we could infer that behaviour of such key people may be influenced by Trump's Twitter activity, resulting in probable flow of information and potential market influence.

Key finance & business players:

The list of financial accounts set up manually for this study (based on the list from Business Insider) contained business accounts such as Deutsche Bank, Goldman Sachs, Blackrock, Bloomberg Markets, Bloomberg Business, JP Morgan, Morgan Stanley, Bank of America, Federal Reserve, New York Stock Exchange, as well as individual accounts such as Warren Buffett, Mike Bloomberg, and other key people in the financial industry.

Our purpose here in deciding these entities were influential and prominent enough, and well representative of the business and finance sectors, was to be able to use them as the measure of whether, more broadly, these sectors are well represented in and make up significant proportions of Trumps Twitter community. Our reasoning here being that if we saw a large portion of these entities in closely linked communities to Trump, we could infer they were more likely to act as proponents to (retweeting or tweeting about) his market sensitive statements.

We also believed it reasonable to assume that large numbers key financial players close to Trump and within his community, as deemed by the community analytics tools studied in COSC2671, it could mean these significant market players were acutely attuned to Trump's Twitter activity, were actively monitoring it, and in turn using them directly as markers in their daily business decision making, such as trading up or down particular stocks, and in turn acting as the guide to the rest of the market investors for down the line, less close to the source, i.e. mum and dad investors. A conclusion for cause for the fluctuations could then perhaps be drawn as the activity of multiple key players 1-degree separated from Trump as the cause, not solely Trump himself.

2nd degree friends:

Two additional, yet unsuccessful attempts were made to retrieve data that we believed could provide detailed insights into the interaction of Trump and his online Twitter community.

1. Obtaining the tweets authored by his followers to gain insights into their interactions in the form of retweeting and liking was not possible due to time access limitation of the Twitter API.
2. Developing a comprehensive list of his second-degree followers (the followers of his followers), to build more convincing community and influence flow models was not possible either due to the limitations of Twitter API.

We believe both these additional datasets would have significantly improved the analysis and conclusions both in Part A and B of this research.

We, however, were able to extract a comprehensive list of Trump's second-degree friends (i.e. the accounts followed by the 47 friends that Trump follows). This became a much easier task for the API to complete as it was dealing with only 47 accounts, each of which were only following a small number of other accounts – the total number of accounts retrieved was approximately 43,000.

Our purpose for extracting this list of 2nd degree friends was less about testing an unique hypothesis, and more an attempt at adding additional data to our community analysis modelling to compensate for the small number of immediate 1st-degree connections we were limited to extracting from Trump's account.

To summarise, the data retrieved for the network analysis contained the following:
- a 'followers' sample of 10,000 accounts,
- all the first and second-degree friends, -- describe this better
- a sample of 3,000 tweets,
- the first-degree user mentions and replies from those 3,000 tweets,
- and the second-degree user mentions and replies from the last 100 tweets of each account that Trump mentioned or replied to in the 3,000 tweets.

Data pre-processing:

Standard pre-processing steps were performed mostly while reading and writing Twitter accounts and data extractions to and from CSV files. These included checks for nullity or duplicates and converting to different data types for ease of access. The details of these, however, are not relevant for the investigation.

A lot of the pre-processing steps were not required as the data extracted from Twitter, was by design already in suitable and uniform format (i.e. already functional, a of a specified format within Twitter). All the tweets had authors, the text in the tweets all obeyed the designed constraints of the twitter program, etc.

The data for the 'user mentions and replies' network graphs (i.e. the graph outlining who Trump directly interacts with using user mentions and replies) was pre-processed similarly to the friends and followers' graph, ensuring each account was identified by their Twitter screen name, also known as Twitter handle.

To identify each account, the unique Twitter screen name of the account was used. This allowed for comparison against the list of financial accounts to evaluate if any of Trump's followers or first and second-degree friends were financial key people.

Moreover, by using the unique screen name to represent each account, it was possible to compare the account to be added to the network against the already existing nodes in the network, to ensure no duplicates.

Data Exploration & Analysis:

The intention of Part A of this research was to explore the interactions between Trump and the immediate communities which exist near and have formed around him in the Twittersphere. Our pursuit for insights is guided by our original assumption that Trumps influence on financial markets via Twitter is made possible by the paths his commentary flows through, and the communities in which it flows to and is considered important by.

This investigation is one based entirely on experimentation, taking the techniques and tools gathered in our studies, coupled with what we believe to be reasoned hypothesise and expectations, and applying them in an almost trial and error approach, to see if any of them yielded conclusive, or at the very least, informative results on whether Trumps influence on the financial markets can indeed be detected and measured via his Twitter interactions.

Exploration 1 – Trumps 1st degree network analysis

Our first approach was to explore through visualisation, the extracted, random list of 10,000 followers and the 1st degree friends that Trump followers – call these collectively a sample of Trump's 1st degree connections – to see whether they form any communities amongst themselves of any significance, and if so, what is that significance. We hypothesised that any community formed would give us confidence that the data we had was useful, and could

reveal, upon further analysis, information regarding their common interests, any other key persons of interest and influence, and potential insights into the flow of Trumps influence.

The network graph of Trump's 1st degree connections (Graph 1) was plotted using the **networkx package**, utilising the **amada kawai** layout. Graph 1, ultimately yielded no useful information. The large number of nodes resulted in incoherence and this method resulted in no formation of communities. Based off this result, we employed the **Gephi** platform to better assist our visual exploration by way of better styling, expanding and changing the colours of the nodes.
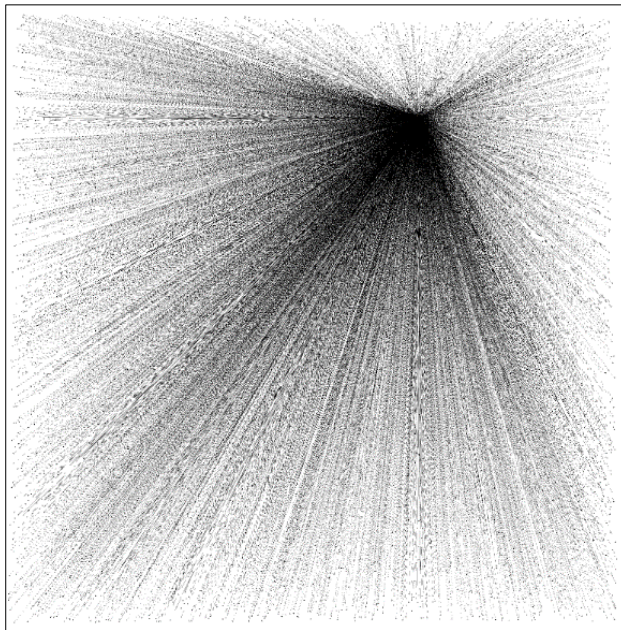
Graph 1: Trump's first-degree connections – the 47 accounts that he followers (termed 'Friends') and the 10,000-account sample of his total 65 million followers.

Beyond not being able to extract any useful conclusions of any sort form Graph 1, flows the obvious conclusion, that it therefor does not provide any insight into whether Trump's network contains any key players from the financial sector – ultimately the type of community kingpins we are hopeful to find in order to support our overarching research quest.

To resolve the concern that we had indeed extracted 10,000 followers, 3,000 tweets, and 42,000 accounts of 2nd connections without any of them containing any of the key financial players that we deemed as gauges of potential market influence, our next step was manual. The individual datasets (in CSV format) containing the friends and the sample 10,000 followers were manually searched to see if any of the accounts matched those accounts in our manually aggregated 'key financial players' accounts dataset, that is, if any financial accounts are among the list of 10,000 followers or 47 friends. Unfortunately, this resulted in zero matches, in either of the datasets. Statistically, given that 10,000 accounts from a possible 65million is an insignificant 0.015%, this could be argued as the expected result, however this finding was disheartening and gave us concern as to whether our research would yield any results of significance at all.

Exploration 2 - community analysis:

Intuitively, we understood that in order to visualise and analyse the communities within Donald Trump's network of followers, it was necessary to obtain more detailed information about the 10,000 followers we had in our study sample, such has who else they followed, and who followed them. Unfortunately, we were again prevented from extracted this imperative data by the Twitter API.

As a result we shifted the community analysis investigation towards an area we knew we could obtain more detailed data, which we hoped would still yield valuable insights - Trump's first and second-degree friends (i.e. who he followed, and who they in turn followed), as well as the first and second-degree accounts Trump interacts with via user mentions and replies (i.e. the accounts he interacts with and the accounts each of them also interact with).
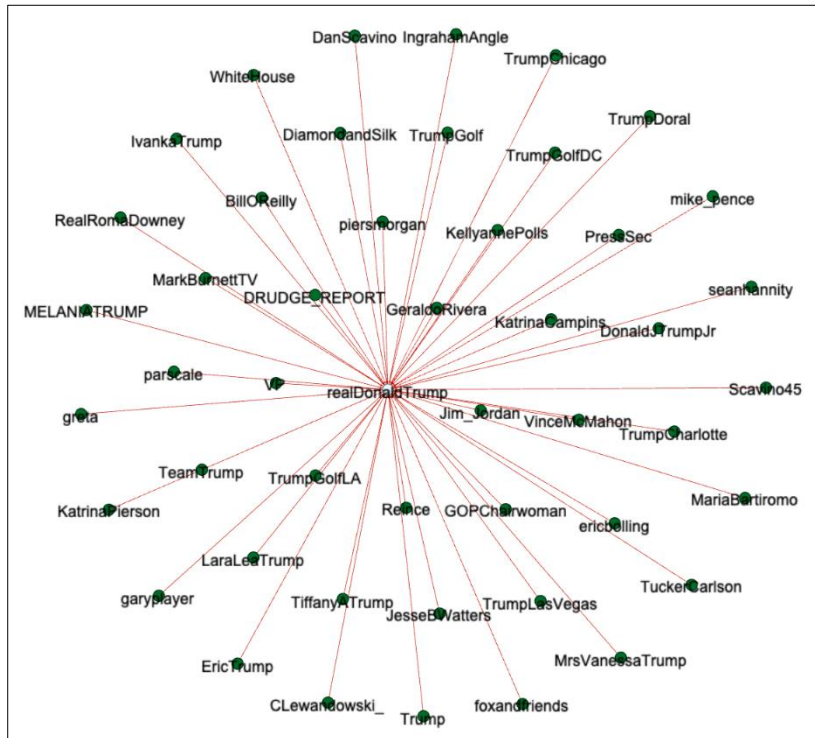

Graph 2 shows Trump's first-degree friends (47 in total).

As can be seen in Graph 2, the limited number of first-degree friends did not seem to form any communities as none of the friends were connected by any edges, a surprising but not overly consequential finding in context of our broader research question, as none of these 47 are considered key players of the financial sectors.

Extending this exploration of communities within Trump's Twitter network, a network graph containing the first **and** second-degree friends, and the connections between them, was constructed, the resulting plot is Graph 3. 14 major communities formed as a result of this analysis - Graph 4 breaks down the relative size of each of the major communities, which is useful in determining which could be the most effective or useful in disseminating Trump's statements.
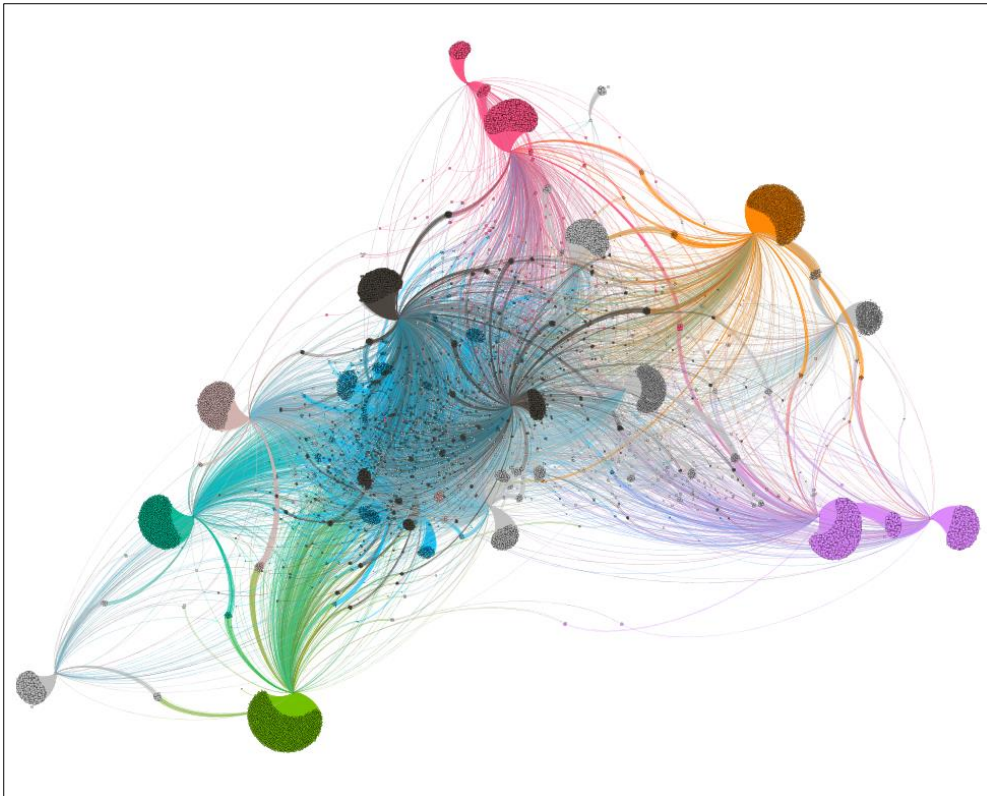
Graph 3: A community detection analysis of Trumps 1st and 2nd degree friendships on Twitter. Constructed using NetworkX package, visualised and partitioned through Gephi platform, where the centrality calculations were omitted and the layout used was Force Atlas 2.
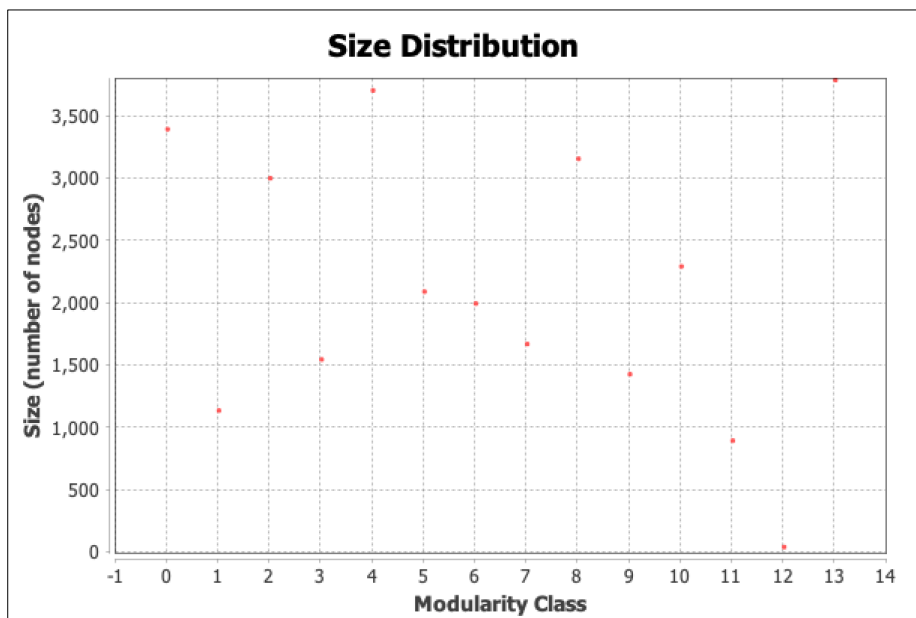

Graph 4: Size distribution of the 14 major communities (modularity classes) formed in the 1st and 2nd degree friendship network analysis of Donald Trump's Twitter account.

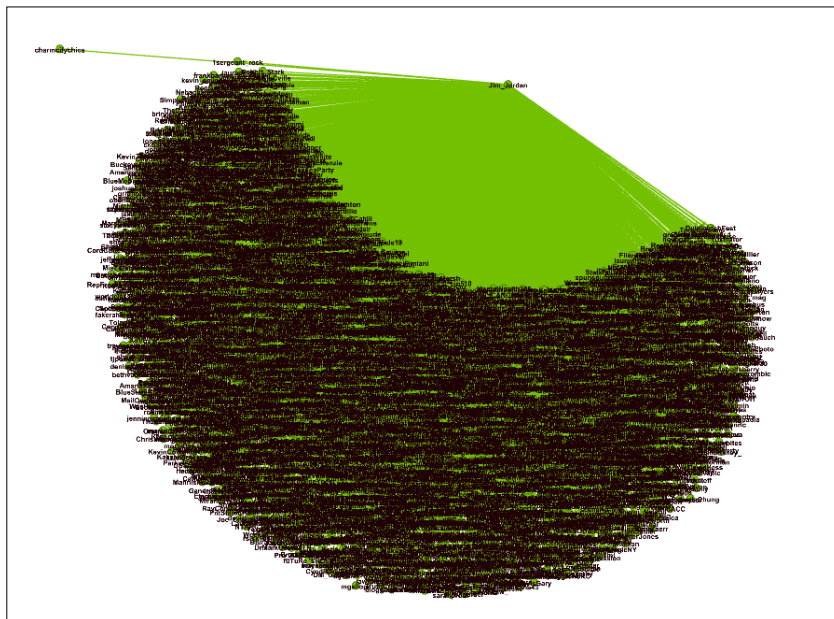As it can be observed in Graph 4, the largest communities formed were "13", "4", and "0" - each containing approximately 3,500 nodes. The next step for us was to continue to explore these communities in hope that what or who they were clustering around would be key financial stakeholders, giving some element of credence to our hypothesis that interaction (pathways and communities) could help our conclusion of market fluctuation can be tied back to Trump.
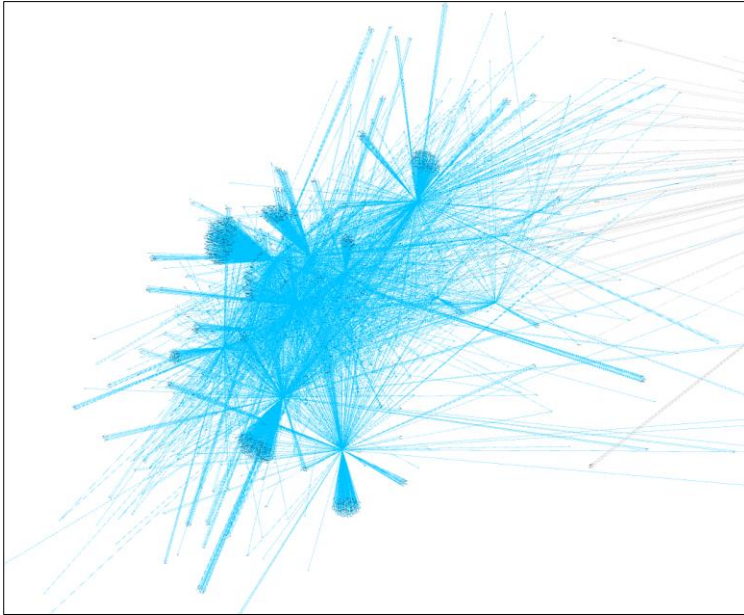
Graphs 5, 6 and 7, respectively show the main community cluster for the three largest of the 14 communities – 13, 4, and 0.

Graph 5: Community 13 is centred around account "garyplayer" and "TrumpDoral".
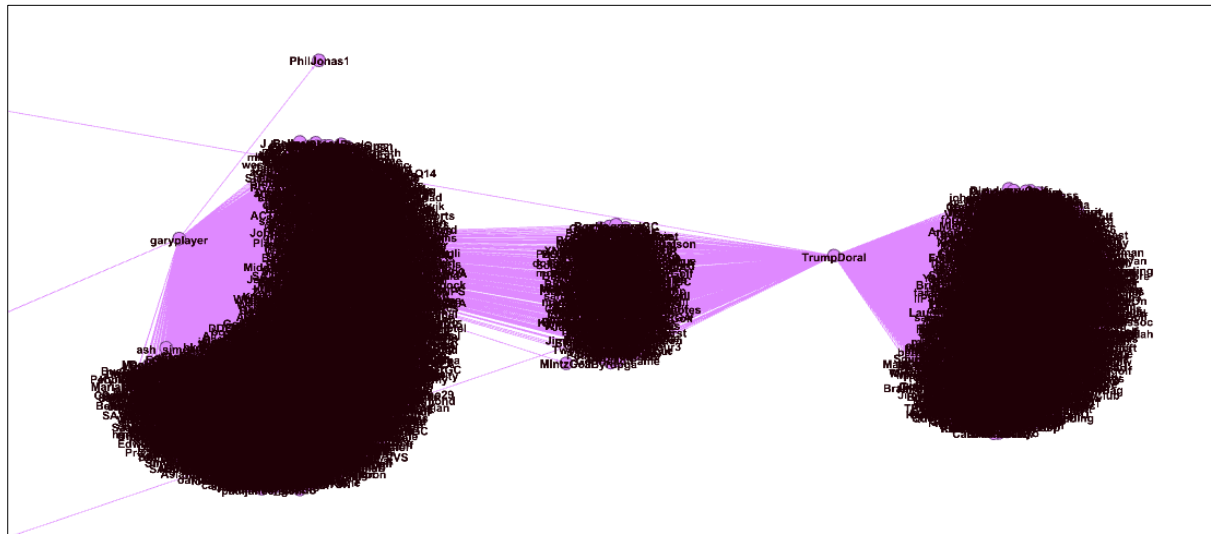
Graph 6: Community 4 is centred around account "Jim_Jordan".

It can be observed, neither of Trump's three largest communities are formed around financial accounts. Upon further investigation, it was observed that none of the 14 major communities were formed around any of the financial accounts we established at the beginning of our investigation. Moreover, by repeating the manual process we conducted in 'Exploration 1' of manually filtering the datasets for any of the 150 financial account we established as 'key players', we found none of his first-degree friends were financial accounts. Interestingly however, 139 of the 150 financial players were represented while filtering the whole list of 42,000 second-degree friends. This finding is significant as it tells us that almost all these players are only two degrees away from Trump – i.e. the content Trump tweets has a fairly straightforward path to, and potentially, through these players. Building from there, perhaps a more interesting finding here, however, is that none of these players seemingly cause a community clustering with any of the other 42,000 accounts. Perhaps a telling result, suggesting that if they do not cause a clustering - i.e. if they are not at the centre of one of the major 14 communities - that the fluctuation of the markets is not caused by flow of the commentary through them…they, perhaps, are potentially not to blame as much as we predicted.

To investigate this, namely to which community the 139 financial accounts did belong to, the data in the Gephi Data Laboratory, which we were using to form our community's visualisations, was exported to a CSV and filtered for each of the financial accounts. The result was the following dictionary:

Community (modularity class)	Number of Key Financial players
0	65
7	4
8	8
6	2

Table 1: A distribution of the number of key financial players found across the communities. NB: The reason why the values in the dictionary do not add up to 139 is because the dictionary has a set of keys, removing any duplicates. On the other hand, the list of 139 financial accounts does have duplicates in it, as more than one of Trump's first-degree friend can be friends with the same account.

This suggests that the community in which the largest number of financial accounts was present was community 0, containing 65 of our pre-determined key player financial accounts. Based on this we took a closer look at community 0: the graph was filtered to ensure only the nodes part of community 0 remained, the layout was changed to Yifan Hu, and the betweenness centrality was computed using Gephi's Average Path Length statistics to evaluate the centrality of each node and therefore establish the influential nodes in the community. Graph 8 represents our Gephi community visualisation.
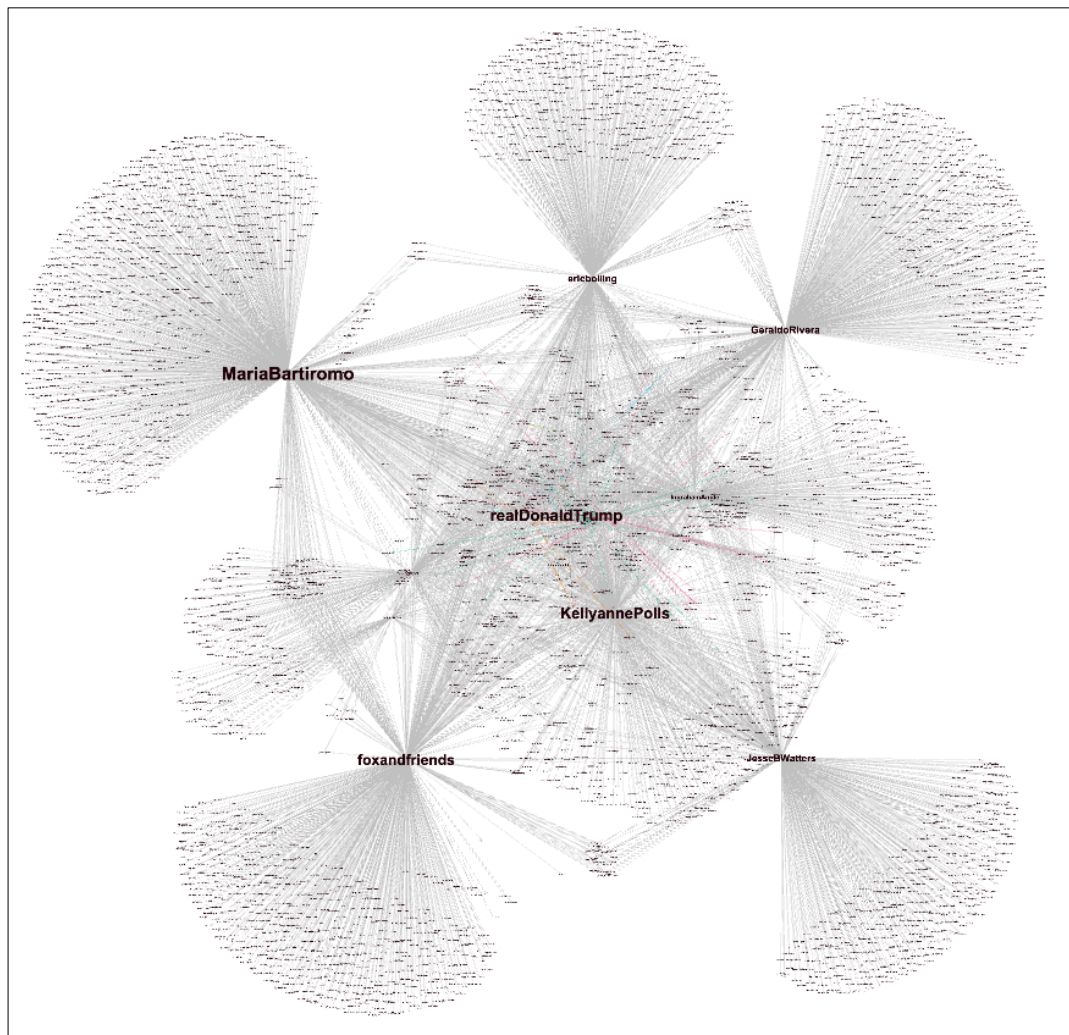
Graph 8: A detailed look at community 0, the community of Donald Trump's Twitter friendships that contains the most financial key players.

In Graph 8, the node size and label is proportional to the betweenness centrality of the node. Noticeably, the most central nodes are Trump, Kellyanne Conway (KellyannePolls, a key Trump advisor), Fox and Friends (Conservative news program), Geraldo Rivera (attorney), Maria Bartiromo (journalist), Eric Bolling (TV personality), and Jesse B Watters (political commentator for Fox and Friends). From Graph 8 it can be observed that none of the financial accounts we expected to appear, such as Bloomberg markets, Mike Bloomberg, NYSE, or the Federal Reserve, were among the central nodes of this community. However, further analysis has shown that the financial accounts in this community are mostly

followed by Maria Bartiromo, a TV journalist who reports on the American financial markets on her show Wall Street Week [2], making it logical for her to follow key financial people. This perhaps is the flow of commentary that we have been seeking – from Trump to Maria Bartiromo to ~65 of the most influential financial players (not to mention her journalism reporting channels) to the broader community of mum and dad investors?

One final investigation was undertaken as part of our Part A analysis, and that was to observe whether a community analysis of the accounts that Donald Trump most frequently mentioned and/or replied to would produce closer more telling results than the previous community profiling. Our process included using the accounts extracted in the most common user mentions and replies from Trump's tweets and building out a community model in the same manner as we previously did. The accounts were represented as nodes and is depicted in Graph 9.
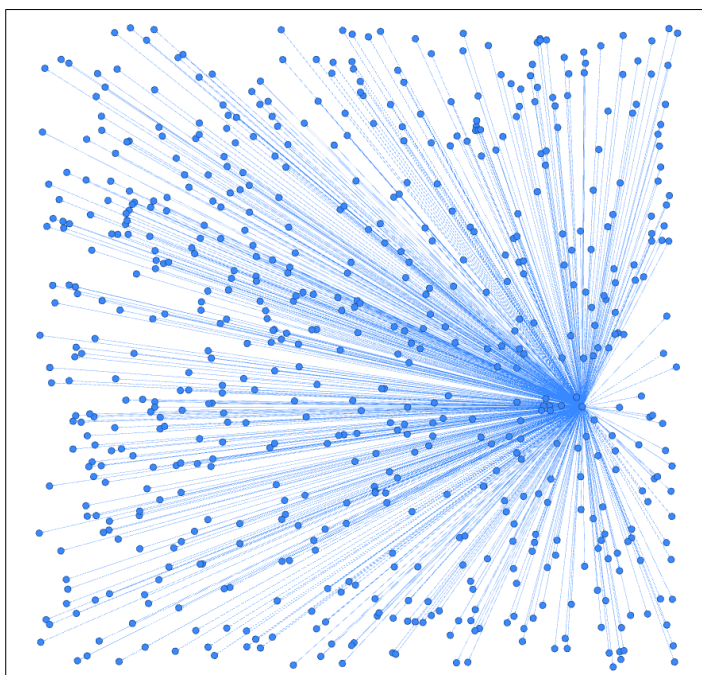

Graph 9: A community model of the account was frequently mentioned by and replied to by Donald Trump on Twitter.

Similar to the friends and followers network, no communities or node centralities could be computed on the first-degree mentions and replies as there seemingly were no other connections between each of them, the only common node was @RealDonalTrump. Hence, for each account that Trump replies to or mentions in his tweets, the last 100 tweets were collected and the user mentions and replies from the 100 sample tweets were added to the network graph. The result was a network containing Trump's first and second-degree user mentions and replies, shown below.

Similar to the friends and followers' network, no communities or node centralities could be computed on the first-degree mentions and replies as there are no other connections than from Trump to the accounts mentioned. In order to explore this further, a random sample of 100 accounts from those that Trump replied to or mentions in his tweets was taken, and for each of those accounts, the last 100 tweets were collected and the user mentions and replies in those tweets were added to the network graph. Graph 10 represent the resulting network graph, containing Trump's first and second-degree user mentions accounts and reply account, on which the Eigenvector centrality was computed using the networkx package
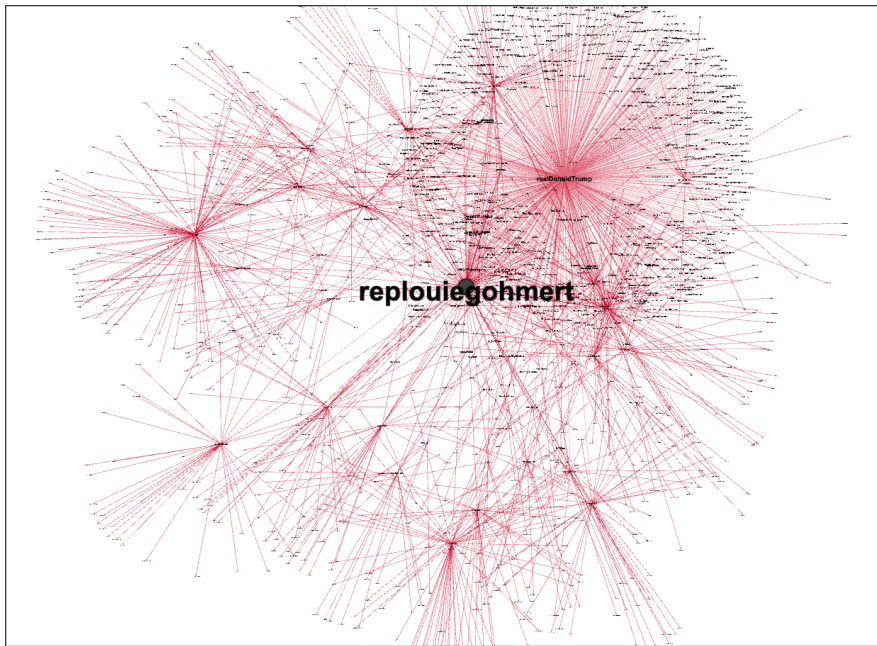
Graph 10: It can be observed that U.S. Representative Louie Gohmert seems to be the node with the highest Eigenvector centrality, suggesting that important and influential users have mentioned him in their tweets or replied to him.

On the other hand, computing Betweenness Centrality using Gephi for the same graph outputted the following result, in which it seems that Donald Trump is at the centre of interactions (refer to Graph 11).
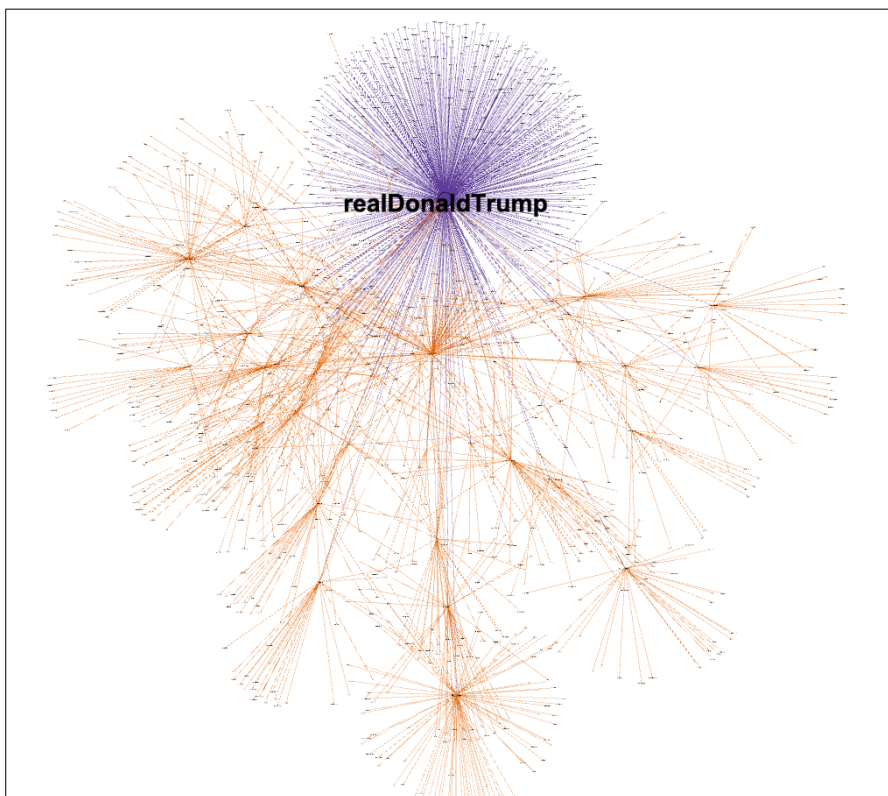
Graph 11: The result of the betweenness centrality seems rather obvious, as Donald Trump is the 'node connecting many of the other nodes in the graph'.

Part A - Results:

The interactions investigation showed that none of Donald Trump's 10,000 followers and first-degree friends were financial accounts, and that most of his communities were formed around his circle of real-life political connections. We did however uncover that 139 of the ~150 key financial accounts were found to be within 2-degrees of separation to Donald Trump. This is significant in that is shows that at the very least, by proximity, that the key decision makers that we likely take our financial decisions guidance from, are close to the source – being DT himself. What it doesn't allow us to empirically conclude however, is if in fact they actually act because of Trump's tweets and more importantly, do they send that influence down the chain to Australia.

Also, despite there being no financial accounts in the 10,000 followers sample, upon investigating the Twitter website it was found that at least 3 financial accounts were amongst Donald Trump's 65 million followers. The reason behind this is the small sample of followers gathered for this study, which represents 0.015% of the total number of followers. As the rest of 99.985% followers were not retrieved using the Twitter API due to its limitations, the website was utilised to evaluate if any financial key people follow Donald Trump, and are aware of the content of his tweets.

Part B – The Action:

Purpose:

The purpose of Part B was to extend on from the interactions around Donald Trump's tweets, into what we call the 'actions' – an examination of the actual content of the tweets for potential market-moving material. In this section we focus on the tweets from Donald Trump, and from Commsec (a major sharemarket broker of the listed companies in Australia), investigating the similarities in content, sentiment of each, event detection flags and any other apparent indicators of the former effecting the latter. We set out with the belief that if Trump's tweets were indeed causing the market fluctuations felt in Australia over the past 6 months, than we would be able to detect the content and topics that caused the fluctuations, or at the very least, isolate the times the fluctuations were caused, by way of event detection analysis.

Data Retrieval & pre-processing:

The data used in Part B was retrieved and pre-processed as described in Part A. The data used through Part B, was restricted to 3,000 tweets from Donald Trump and 3,000 tweets from Commsec over the course of 6 month: April to September 2019.

Data Exploration & Analysis:

Exploration 1 – Text Analysis:

We start off from the most basic level of analysis and count the most frequent terms used in Tweets, collected from both Commsec and Donald Trump. Commsec being an Australian share trading platform uses the term "us" the most, which refers to the United States. This is not surprising given the U.S. is the biggest economic power in the world. Other terms like "aussie", "market", "close", "shares" and "profit" confirms that the performance of the domestic and international financial markets is central to this platform. On the other hand, Trump's most frequent terms are much broader in context, and are not focused on markets specifically, and certainly not only finance or business specific. Within Trump's most frequent tweet keywords, we see terms like "great", "president", "democrats", and "China" – representing a much more political flavour than finance and business. Other terms like "Mexico", "border", "Carolina", "mueller" suggest domestic issues like the wall that Donald Trump wants to build dividing Mexico, and also Cyclone Dorian which has recently hit North Carolina.

Our analysis determined that there were very few terms that were common between the 3,000 Trump and 3,000 Commsec tweet datasets. The two most commonly frequent terms used by our two key entities were "new" and "us". We determine that "us" relates to the United States but "new" could relate to new tariffs, new wall or new five – decade low

unemployment? Ultimately, neither term is particularly insightful for our overall investigation intent.

Next, we focused our attention on an analysis of the common and frequency of the hashtags used within both tweet sample datasets. A hashtag is a word or phrase used in social media to identify messages on specific topics. It helps identify and group tweets and commentary into certain topics. The top four hashtags used in Commsec tweets are very ASX specific:

- **ausbiz** (Australian Business),
- **rba** (Reserve Bank of Australia),
- **ausecon** (Australia Economy)
- **audusd** (Australia / US exchange rate)

Trump's hashtags most commonly relate to Cyclone Dorian which ranked first on his most frequent. #MakeAmericaGreatAgain or #MAGA and #USMA (The US – Mexico Agreement) were also high on the list – again less financial, more political in nature.

Rank	Commsec Word	Freq	Trump Word	Freq	Commsec Hashtag	Freq	Trump Hashtag	Freq	Commsec Mentions	Freq	Trump Mentions	Freq
1	us	664	great	368	ausbiz	576	dorian	33	CommSec	31	realDonaldTrump	274
2	market	579	president	293	asx	262	maga	15	RBAInfo	18	WhiteHouse	97
3	podcast	532	trump	191	ausecon	200	usmca	12	Reuters	17	FoxNews	45
4	listen	448	thank	185	markets	96	kag2020	8	economics	9	TomFitton	36
5	video	438	democrats	182	gold	45	g7biarritz	5	WiseTechGlobal	8	GOPChairwoman	32
6	report	430	new	146	rba	43	socialmediasummit	5	MontakaGlobal	8	VP	32
7	ios	409	news	138	audusd	34	salutetoamerica	5	BloombergAU	8	NHC_Atlantic	31
8	shares	402	people	136	fx	32	nc09	4	CBAnewsroom	8	JudicialWatch	30
9	aussie	365	china	118	rates	31	dday75thanniversary	4	CNBC	8	foxandfriends	29
10	close	310	country	117	bonds	28	potusinjapan	4	BubsAustralia	8	DonaldJTrumpJr	29
11	year	298	today	108	aud	27	nc03	3	whispir	8	POTUS	27
12	executive	283	media	105	usdaud	27	fakenewscnn	3	business	8	Jim_Jordan	27
13	trade	276	border	104	asx200	27	hurricanedorian	3	apngroup	7	RNCResearch	22
14	series	266	fake	103	ironore	25	spygate	3	KathmanduGear	7	IvankaTrump	22
15	stocks	246	big	102	ozearnings	24	trump2020	3	marleyspoonau	7	FLOTUS	21
16	rose	221	united	98	oil	20	dday75	3	amaysimAU	7	RepMarkMeadows	20
17	fell	208	would	96	stoxx600	17	maga2020	3	NanosonicsLtd	7	usminority	19
18	profit	206	america	93	ftse100	17	irgc	3	medibank	7	CNN	18
19	results	194	years	87	dax	17	usstatevisit	3	IAGAust	7	MariaBartiromo	17
20	day	188	one	87	spi	16	keepamericagreat	2	ETFSecuritiesAU	7	TeamTrump	17
21	pts	188	states	84	auspol	14	unga	2	AFP	6	mike_pence	17
22	fy19	181	mueller	81	cac40	12	trumprally	2	Kogan	6	seanhannity	15
23	morning	175	like	80	interestrates	7	usmcanow	2	Transurban	6	LindseyGrahamSC	14
24	asx	175	never	79	breaking	7	wgdp	2	Mirvac	6	senatemajldr	13
25	group	173	american	75	lithium	5	hurricane	2	ecofibre	6	CLewandowski_	13
26	week	172	much	72	retail	4	g7summit	2	ReutersBiz	5	GOPLeader	11
27	ceo	170	many	71	gdp	4	israel	2	Data3Limited	3	SteveScalise	11
28	ltd	165	u	69	nzpol	3	america	2	corelogicau	3	LouDobbs	11
29	result	160	state	68	fintech	3	kentucky	2	ReserveBankofNZ	3	nytimes	11
30	higher	157	north	66	energy	3	pledgetoamericasworker	2	ASX	3	AbeShinzo	11
31	full	149	good	65	crude	2	womenfortrump	2	baseresources	3	RepDougCollins	10
32	mid-session	147	house	63	infantformula	2	walkaway	2	sunriseon7	3	BreitbartNews	10
33	speaks	133	want	63	agl	2	barry	2	alcidion	3	IngrahamAngle	9
34	price	131	carolina	63	imf	2	icymi	2	IMFNews	3	RudyGiuliani	9
35	economic	130	mexico	63	weo	2	boycottcnn	2	ABSStats	2	EricTrump	9
36	points	130	get	62	rbnz	1	july4th	2	craigjamesOZ	2	KimStrassel	9
37	local	126	time	62	metals	1	g20osakasummit	2	CNBCi	2	GOP	9
38	gains	121	going	61	retailers	1	orlando	2	CNBCnow	2	KTHopkins	9
39	company	119	done	60	etfs	1	americafirst	2	tictoc	2	DanScavino	9
40	business	119	job	59	smsfs	1	fakewhistleblower	1	Scutty	2	FoxBusiness	8
41	best	118	deal	58	fomc	1	ukrainescandal	1	ReutersGMF	2	marklevinshow	8
42	net	118	us	57	smsf	1	rt	1	AP	2	PressSec	8
43	new	114	back	56	dairy	1	hispanicheritagemonth	1	ANZ_Research	2	ByronYork	8
44	revenue	112	u	55	wealtheffect	1	fakenews	1	roymorganonline	2	Franklin_Graham	8
45	record	111			ruokday	1	nationalvoterregistrationc	1	realDonaldTrump	2	JesseBWatters	7
46	earnings	111			trustthesigns	1	hbdusaf	1	markets	2	MorningsMaria	7
47	lift	109			ruok	1	leadright	1	fastFT	2	Scavino45	6
48	worst	108			tradewar	1	fisa	1	MarketWatch	2	parscale	6
49	b	107			us	1	aiingovsummit	1	ftfinancenews	1	USCG	6
50	near	105			china	1	video	1	CNBCClosingBell	1	fema	6

Common References

Table 2: Top 50 most frequent terms, hashtags, and user mentions in the 3,000 Commsec and Donald Trump tweet datasets.

Finally, we analyse the user mentions which are a way of notifying other users of the tweets you are posting but specifically mentioning their account name within your post. It is not surprising that the top references that Commsec mention are some of the major economic entities in Australia like RBAinfo, Reuters, BloombergAu and BubsAustralia. These are predominantly research entities and news platforms. This is not too dissimiliar to Trump's user mention which also include a number of news platforms like Foxnews and JudicialWatch. Tom Fitton who is also mentioned, is the president of JudicialWatch. This is centred around the potential impeachment of Trump, which has been the trending news story out of the U.S. recently. We also noted that the user mention of @realDonaldTrump (Donald Trump's twitter handle) was found to be a user mention by both Trumps own account, and Commsec. It however did not rank highly in Commsec's list of user mentions, suggesting that he may not be having as much direct influence to the average mum and dad traders, via the Commsec Twitter account as much as we first thought.

This analysis gives us insight into the content that is being tweeted. From this we know that Commsec tweets on more specific content focused on the Australian stockmarket. Donald Trump specifically tweets on more political matters focusing on the U.S. Events such as Cyclone Dorian, potential impeachment and the Mexican wall seem to be the themes that we can correlate between our analysis and the big news stories coming out of U.S in the last couple of months.
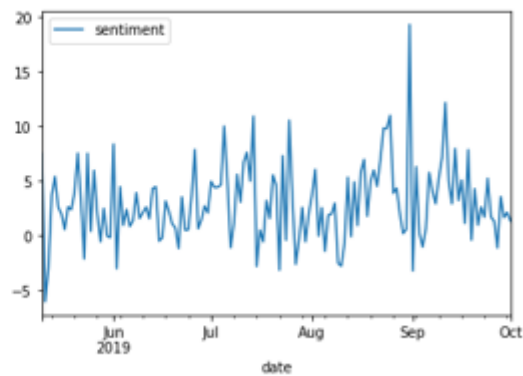
Exploration 2 – Sentiment Analysis:

In this phase of our investigation we focus our analysis on whether they are positive or negative sentiments tweeted. Our hypothesis is that the tone and sentiment of Trump's tweets could present correlations with the fluctuations of the ASX.

Here we have elected to use the Vader approach which uses a lexicon which has been built from social media. This approach uses features like adjusting the sentiment scores for punctuation marks, capitalisation, degree modifiers like 'super' which infer a greater degree of something and shifts in polarity. This lexicon has been specifically developed for social media and has a sentiment score which ranges from -4 to 4. Below we have summed these scores on a daily basis, as other frequencies produce a lot of noise.
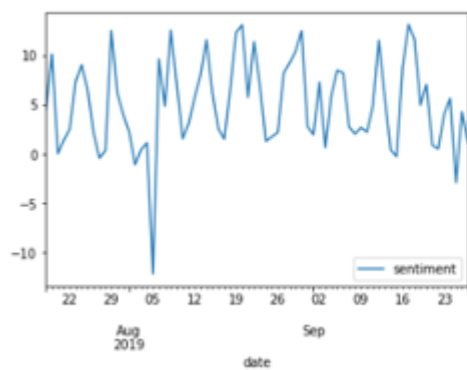
Graph 12: Sentiment analysis for Trump Aug-Sept 2019, and Commsec Jun-Oct 2019.

Note the two different time scales which is an affect of the Twitter API's limitation in allowing us to specify specific time period over which the tweets were created – had we had this capability we would have aligned the two datasets to cover the same period of time.

There appears to be more negativity in the Trump account, as there appears to be more instances when the time series crosses below 0. Eyeballing the two accounts over the same date range there doesn't appear to be a correlation. This would suggest that Trump's sentiment does not dampen the Commsec sentiment. This makes sense as we have established in the previous section as different themes appear to be prominent in the two accounts.

As can be seen in Graph 13, there is a massive negative spike on the 5th of August in the Commsec account which does not register on Trumps account. Conversely, there is a massive positive spike in the trump account just before September. To explore and help understand the apparent lack of causation, we ran manual research on whether there we any significant events at either of those times. We found that on August 5th there was an escalation of the US – China trade war. The stock market fell 192 points in Australia and 90b was wiped off the market. This was the worst one day decline in 1 year. China had retaliated against the latest round of US tariffs by devaluing the yuan to its lowest level than more than a decade.

Graph 13: Screen grab of the ASX200 April to October 2019.

The sentiment analysis for Commsec picks this particular event up but was not detected on Trump's sentiment analysis. We believe this is most likely due to this type of news perceived in Australia as negative, i.e. due to our proximity and reliance on China as a trading partner, whereas the US could perceive with nonchalance or even positive sentiment – i.e. Trump's followers would perceive a trade war with China as a positive event or development for the US whereas the investment community would translate this to poor stock performance and returns.

The charts in Visual 1 give us a further view into to overall sentiment of the two accounts we are trying to draw conclusions between. Trump's tweets fluctuate greatly in terms of positive and negative sentiment, whereas Commsec is more neutral – arguably, the more responsible type of sentiment to maintain given the responsibility it holds in affecting broader investor behaviour.
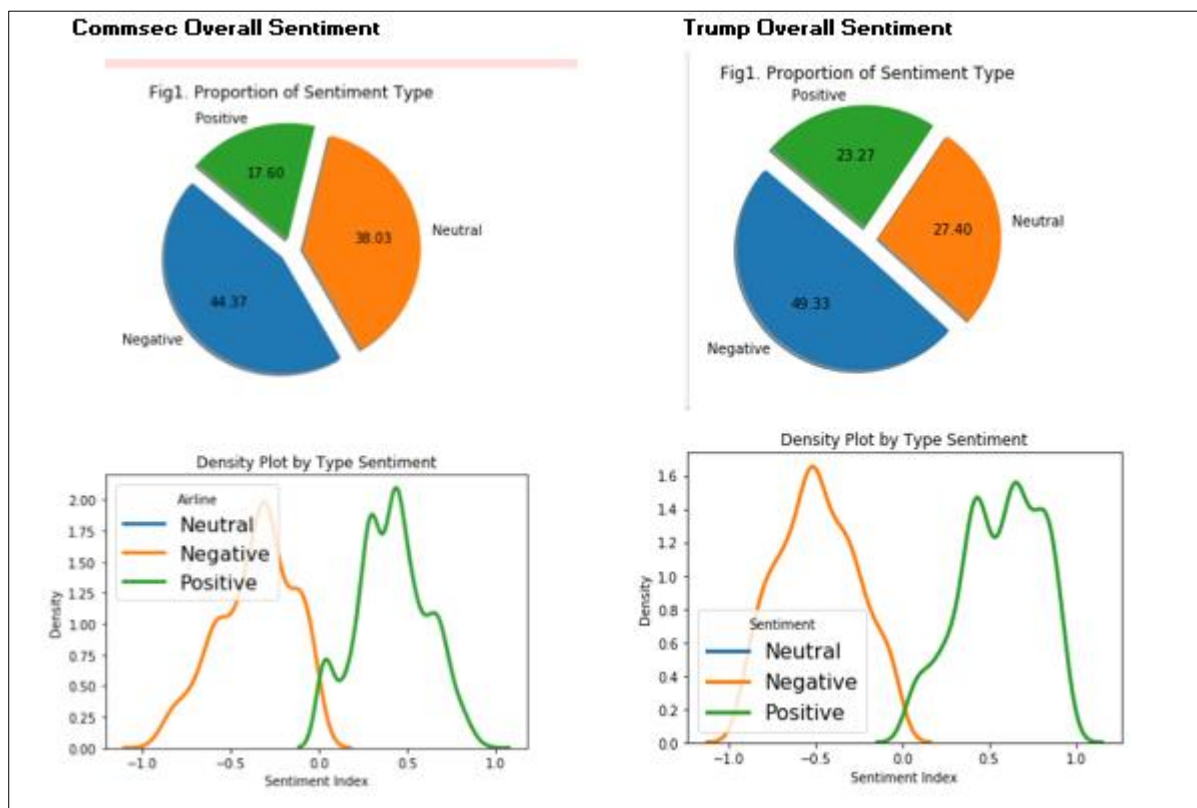
Visual 1: Overall sentiment analysis for Commsec's tweets (left) Trump's tweets (right).

Exploration 3 – Topic Modelling and Evaluation:

Topic Models allow us to explore and summarise the topics and ideas being discussed within the 3,000 tweets for each account. The preliminary analysis shows that we should have different topics for Trump and Commsec, considering all things including politics, geography, economic treaties and agreements, etc. For example topics like Cyclone Dorian which hit North Carolina, Impeachment, Mexico Border, US/China Trade war is what we expected to result from exploring the Trump tweet dataset. For Commsec, we expected more domestic and economic events, and perhaps more observational commentary related topics, as Australia is very rarely the instigator of events, and more so the bystander. Topics related to our domestic markets such as interest rate cuts and fiscal policy, China, property market discourse, and company reporting activity was what we expected to see as most prominent.

Initially we used the default topic number of 10 for performing our topic evaluation.

In Table 2, we can see the 10 topics for Trump account. This works by constructing a document/term from the set of tweets, using a TF-IDF approach. Some of the topics could be considered obvious, for example topic 0 relates to the tariff's imposed on China and the hope of job creation for America. Topic 2 relates to the democrat leader's (Joe Biden) endeavour in getting Trump impeached. Topic 3 is regards to Mueller report which specifies that Trump had called upon Australian prime minister Scott Morrison to assist in discrediting the claim of Russian interference in the US election. Topic 5 relates to the Mexican border and Topic 9 clearly relates to Cyclone Dorian.

```
display_topics(ldaModel, tfFeatureNames, wordNumToDisplay)

Topic 0:
realdonaldtrump united states president whitehouse job tariffs want great tonight
Topic 1:
great news media fake election best federal going world billion
Topic 2:
president trump realdonaldtrump joe like said potus democrats mark biden
Topic 3:
mueller people report right prime minister washington help long believe
Topic 4:
china far way news media clinton iran fake hillary countries
Topic 5:
border years new democrats trump mexico president obama tomfitton congratulations
Topic 6:
economy good bad work real gopchairwoman democrat cnn lost totally
Topic 7:
country american people history jobs things john witch hunt new
Topic 8:
great today america democrats house deal day make trump country
Topic 9:
thank big north realdonaldtrump carolina state really dorian court hurricane
```

Table 2: Top 10 topics which were formed from Trump's tweets.

We explored this visually as seen in Visual 2. By adjusting **lambda,** we get a ranking of terms which appear more specific to the topic, visualised as a greater ratio of red to grey. The Mueller investigation is visualised below as the red circle. The terms "Mueller", "Report", "Prime", "Minister" are terms which are very specific to this topic and you can see the red circle is quite distant from the other circles which reflects this a vastly different topic to the others, where other sub-topics are probably non- existent.
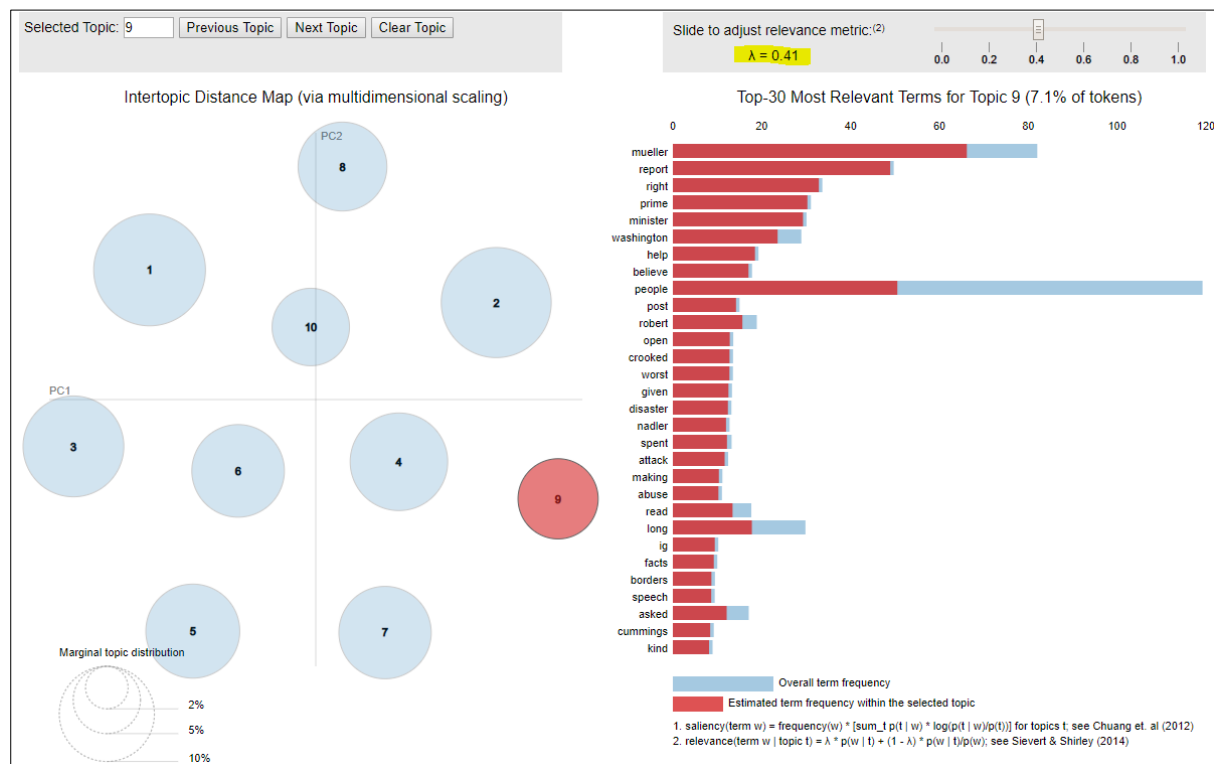
Visual 2: Topic modelling of Trumps tweets.

Running the same analysis for the Commsec set of tweets, we saw that as least one core topic, topic 10 (LDA visual) is Trump related. A major concern for the ASX (particularly Australian mining companies) was the US China Trade War. It can be seen that this topic identifies with terms such as "Trump", "China", "Tariffs" and the companies it would impact "fmg" (Fortescue Metals), "s32" (South 32), and "BHP". Like before lambda was adjusted to weed out the generic terms*.*

This was a significant breakthrough in our research. This LDA visual, is confirmation that Trump's tweets which invariably precede his strategies and objectives, do impact the ASX. However, this is not to a great scale in the period we are analysing.

Other topics clearly identified is the reporting season where companies disclose their financial performance (topic 0). Topic 2 relates to a Commsec run podcast/show on which they have senior executives discuss their Company's performance and prospects. Topic 6 is related to the RBA interest rate decision on the first Tuesday of every month.

```
display_topics(ldaModel, tfFeatureNames, wordNumToDisplay)
```
```
Topic 0:
year profit results fy19 points result net revenue reported video
Topic 1:
ausbiz report ausecon record australia year july june high august
Topic 2:
executive series podcast listen ios ceo video group speaks company
Topic 3:
report sharemarkets ausbiz key trump bhp investors dividend trade update
Topic 4:
commsec china business ausbiz markets market economic latest news podcast
Topic 5:
earnings dow sales jones market nasdaq pts group close gold
Topic 6:
week rate ausbiz rba ahead bank bn podcast unchanged listen
Topic 7:
price rose gold fell pts oil energy trading ounce european
Topic 8:
best day new worst stocks market prices australian performing snapshot
Topic 9:
asx aussie shares market close session trade listen podcast ios
```
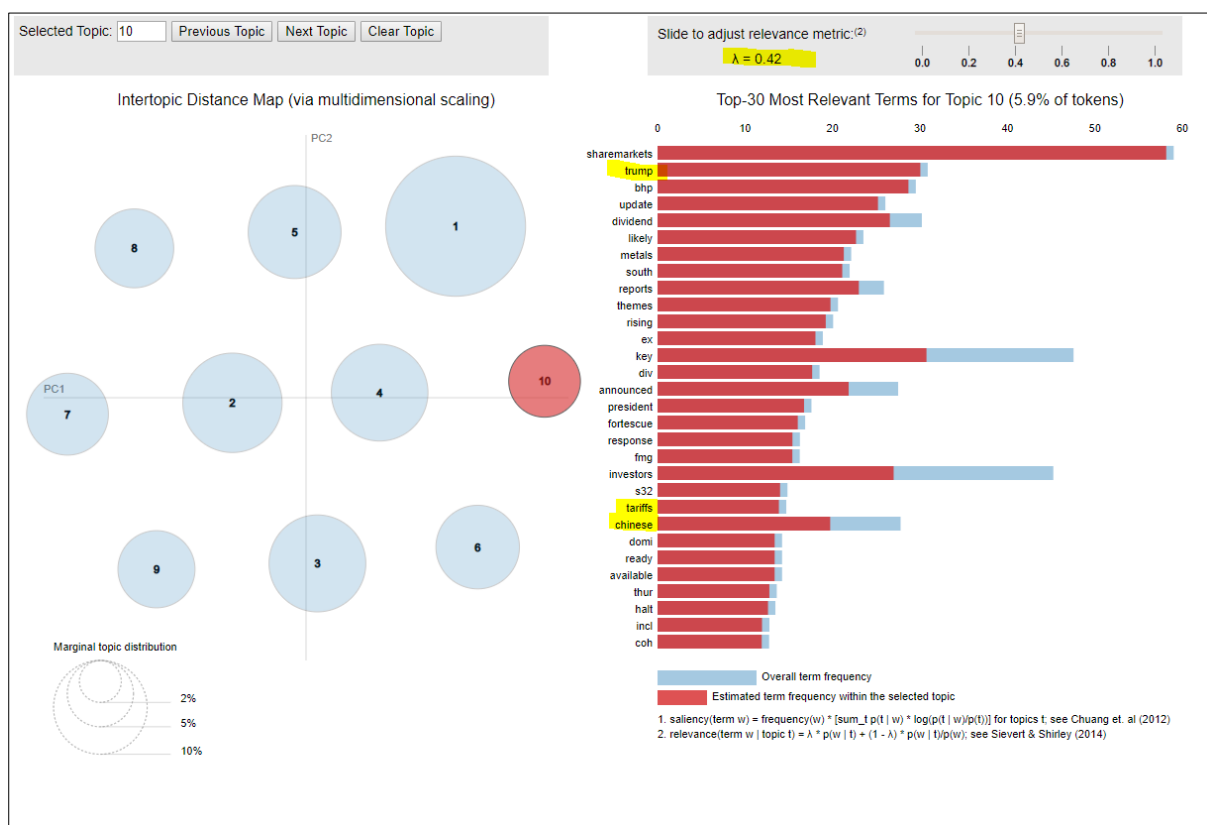
Table 3: Top 10 topics which were formed using Commsec tweets.
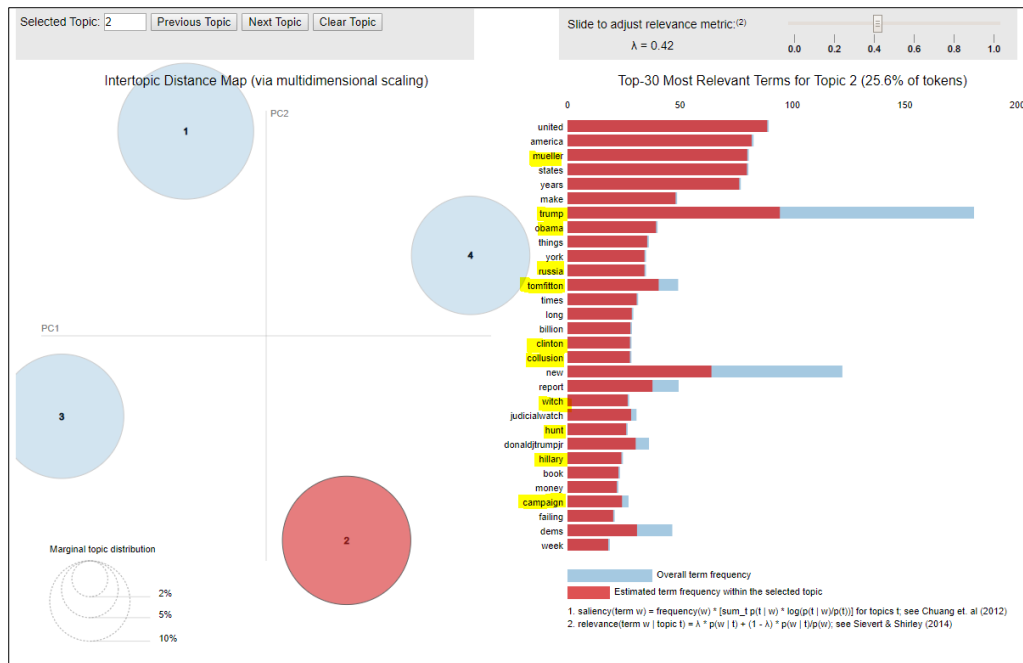

Visual 3: Topic modelling of Commsec tweets.

Topic Evaluation provides an opportunity to further refine our topic modelling. We can utilise a couple of approaches to find the optimal number of topics. Both approaches are suggesting a small number of topics as we want a larger likelihood and or a smaller perplexity score. This is strange and does not make sense as we know Trump is tweeting on a range of topics for the last few months. We decided to experiment with 4 topics.
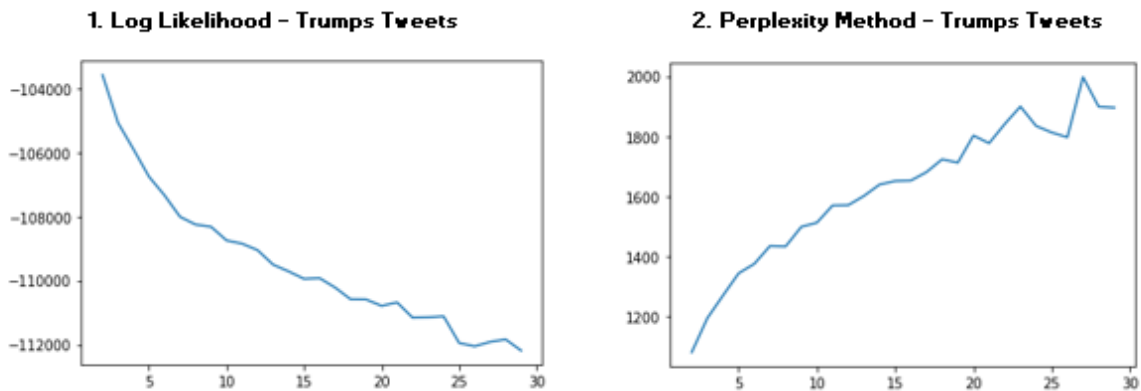

The result of reducing the number of topics in topic analysis can be seen in the below visualisations. The topics are further away from each other reflecting their distinctness and lack of association with other topics. In theory, distinctiveness could provide significant, conclusive evidence that a particular or small group of topics can most commonly be used to categorise the commentary from the source – in our case Donald Trump or Commsec.

In the case of Trump, we were looking for conclusive results that at least one of the major topics centred around something 'Australian' such Asian-Pacific trade, or financial activity – unfortunately no such results emerged. The Mueller Report, along with other key national political topics appear to control most of the conversation.

Visual 4: Revised Topic modelling of Trump tweets using only four topics for analysis.

Reiterating the topic evaluation process for Commsec, using log-likelihood and perplexity methods, produces similar results to the Trump Tweets, in that 4 major topics were identified. Surprisingly, the LDA visual (as seen below) reflects a topic which features "trump", "Chinese", and "inflation" which could relate to the US-China trade war and the financial impact on China due to the newly imposed tariffs by the US.
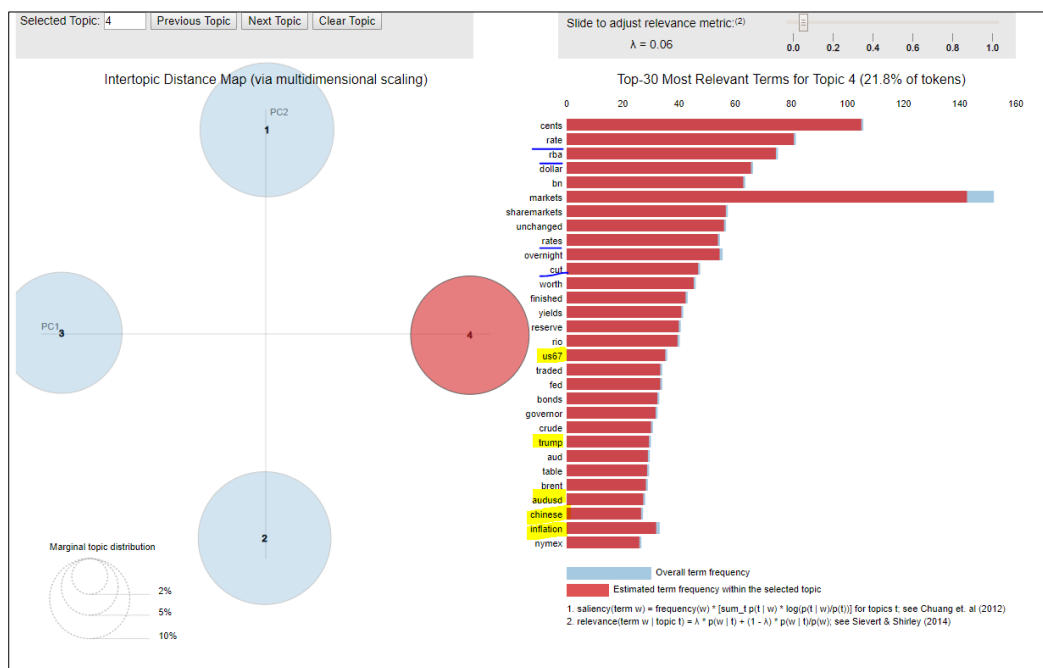
Visual 5: Revised Topic modelling of Commsec tweets using only four topics for analysis.

With interesting results beginning to form in the LDA visual, we extended the analysis by producing wordclouds for the four-key resultant, visual 6 was the result.

Visual 6: wordcloud analysis for Commsec tweets.

Word Embedding is also another technique we used to further explore if there were any clear market sensitive terms and topics used through the Commsec tweet dataset. Word embeddings are useful to show how the key words are semantically related to each other. For example, the words "close", "today", and "worst" are clustered together as they would be linked back to tweets or events (which were tweeted about by Commsec) relating to daily market activity. Similarly, "earnings", "revenue", and "gains" are clustered together as they would normally be tweeted relating to the full year reporting season. Unfortunately, this analysis yielded no useful word associations linking back to our supporting our original hypothesis.
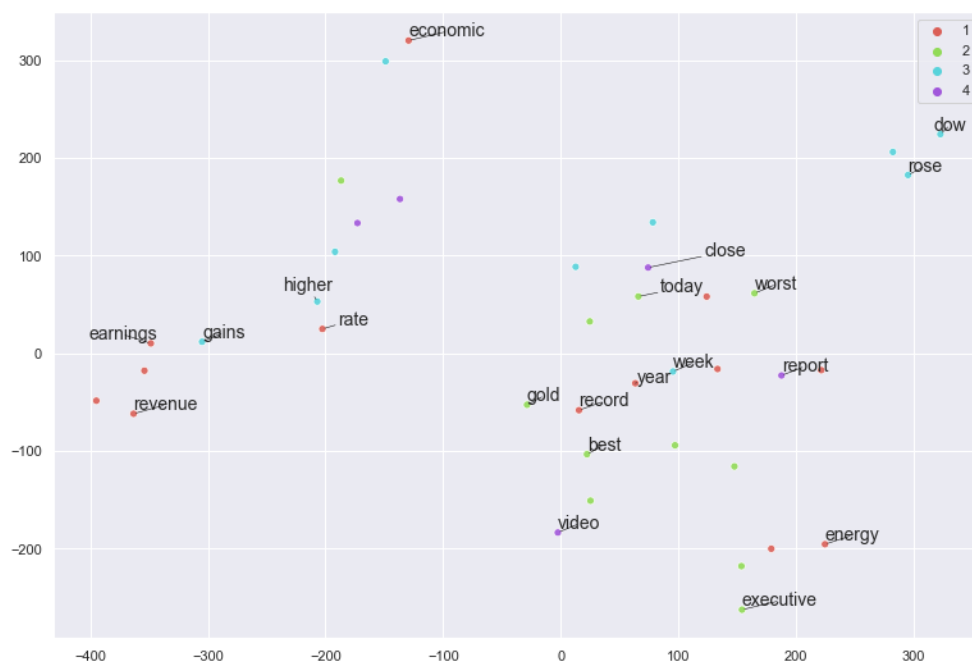
Visual 7: Word Embedding graph, showing semantically linked and associated words from the Commsec tweets dataset.

Part B results:

Similar to Part A, the Action Investigation conducted in Part B resulted in no clear or supporting evidence that activity from Donald Trump and his twitterverse affect the Australian stock market. 3,000 tweets from Donald Trump were analysed for references, sentiment, topics groupings and word associations that could be deemed as Australian specific and therefore market sensitive. Likewise, 3,000 tweets from Commsec, were analysed using the same step by step process. In doing so we were able to pin down, using the LDA modelling process, that words such as "China", "Trump", and "Tariffs" were used alongside significant ASX listed companies (Fortescue Metals, South 32, and BHP). The link here for us (perhaps an attempt to clutch any straw we can), is that investor sentiment and behaviour is often swayed but the performance and activity of the largest market players (here in Australia they do tend to be our mining companies), and if these players are being linked to market upheavals and trade warmongering between Trump, trade agreements and China, then this is a likely channel of influence to the average mum and dad investors of Australia.

Conclusion:

Our endeavours to empirically detect and measure whether Donald Trump's tweets cause the fluctuations in the share prices of companies listed on the Australian Stock Exchange ultimately produced no conclusive results.

The investigation into Trump's interactions via Twitter, and his community of followers and friends, was undertaken to help determine the flow of the influence that we hypothesised his tweets had on the share market in Australia. Using the techniques and tools such as the NetworkX package in python and Gephi, we did not find any clear communities, or interactions between Trump twitter account and other accounts that we could use to conclude any flow of influence. The most useful insight gained from this, Part A, of our research was that the vast majority of the financial key players we used as our gauge were within 2-degrees of Donald Trump (i.e. they were friends of his friends) – 139 out of 149. More interesting still, is that 65 of these 139 were found to form a community dominated by central nodes such as Kellyanne Conway, Fox and Friends, Maria Bartiromo, Eric Bolling, and Jesse B Watters – all highly influential journalists or public personalities. This could suggest that Trump's commentary is amplified and given credibility as it flows outward through these media personnel.

Part B – the investigation of Trump's activity on Twitter as well as Commsec's – also resulted in inclusive evidence that what Trump says via Twitter can be drawn on as a causing the fluctuations. Our research did determine a number of keywords are often used by Commsec such as "China", "Trump", and "Tariffs", and that these used alongside terms relating to our large market stalwarts such as BHP and Fortescue Metals, perhaps indicated that there were links between what Trump commented on and what they themselves made comment on.

Our logic being that investors could be taking their information from Commsec's tweets, and making decisions based on them, causing the fluctuations. However, the results were inclusive on this front, but it does present a potential lead, or starting point if further analysis was undertaken.

Through the entirety of the project, the biggest challenge, and ultimate limiting factor in our research was the relatively small samples of data that we used. Unfortunately, the Twitter API restricted our ability to retrieve large enough datasets to enable us to draw any meaningful conclusions from, and we believe this is the reason we were not able to conclude one way or the other whether the fluctuations were caused by Trump or not.

In saying this, we believe that with more data behind our analysis we would uncover significantly different, and much more conclusive results as to Trump's affects on the ASX performance.

Sources and References:

[1] S. Bartholomeusz, "The Volfefe index shows Trump's tweets come at a cost", The Age. [Online]. Available: https://www.smh.com.au/business/markets/the-volfefe-index-shows-trump-s-tweets-come-at-a-cost-20190911-p52q8n.html. [Accessed 08-Oct-2019].

[2] R. Graham, "The 125 most important finance people you have to follow on Twitter", *Business Insider*. [Online]. Available: https://www.businessinsider.com/people-to-follow-on-twitter-from-finance-2017-6/?r=AU&IR=T. [Accessed: 03- Oct- 2019].