



Trump's COVID-19 tweets and Dr. Fauci's emails

David E. Allen^{1,2,3}  · Michael McAleer^{2,4,5,6,7}

Received: 27 July 2021 / Accepted: 12 December 2021 / Published online: 16 January 2022
© Akadémiai Kiadó, Budapest, Hungary 2022

Abstract

The paper features an analysis of former President Trump's early tweets on COVID-19 in the context of Dr. Fauci's recently revealed email trove. The tweets are analysed using various data mining techniques, including sentiment analysis. These techniques facilitate exploration of content and sentiments within the texts, and their potential implications for the national and international reaction to COVID-19. The data set or corpus includes 159 tweets on COVID-19 that are sourced from the Trump Twitter Archive, running from 24 January 2020 to 2 April 2020. In addition we use Zipf and Mandelbrot's power law to calibrate the extent to which they differ from normal language patterns. A context for the emails is provided by the recently revealed email trove of Dr. Fauci, obtained by Buzzfeed on 1 June 2021 obtained under the Freedom of Information Act.

Keywords Trump · Tweets · Text mining · Sentiment analysis · Word cloud · COVID 19 · Stock market · Dr Fauci emails

JEL Classification C19 · C65 · D79

In Memoriam: Michael McAleer passed away on 8 July 2021, aged 68, after a decade plus struggle against cancer. He was a remarkable scholar with interests in econometrics, economics, financial econometrics, finance, tourism economics, energy economics, the economics of patents, bibliometrics, and more recently statistical applications in COVID-19 research. His research output was phenomenal with some 415 publications and some 6,786 citations indicated on Publons, and 1,270 referenced pieces on Google Scholar with some 23,273 citations. He will be greatly missed by his vast network of global scholars and research collaborators. We contribute to his memory by publishing one of his last co-authored contributions.

✉ David E. Allen
profallen2007@gmail.com

¹ School of Mathematics and Statistics, University of Sydney, Sydney, Australia

² Department of Finance, College of Management, Asia University, Taichung City, Taiwan

³ School of Business and Law, Edith Cowan University, Joondalup, Australia

⁴ Department of Bioinformatics and Medical Engineering, College of Information and Electrical Engineering, Asia University, Taichung City, Taiwan

⁵ Discipline of Business Analytics, University of Sydney Business School, Sydney, Australia

⁶ Econometric Institute, Erasmus School of Economics, Erasmus University Rotterdam, Rotterdam, The Netherlands

⁷ Department of Economic Analysis and ICAE, Complutense University of Madrid, Madrid, Spain

Introduction

Former President Trump is a controversial figure in the media and in political commentary. He made heavy use of his former Twitter account and was always ready to call out 'fake news'. According to Wikipedia, Donald Trump's presence on social media attracted attention worldwide since he joined Twitter in 2009, with the handle @realDonaldTrump, having over 88.9 million followers by 2021, culminating in his banning from Twitter for at least two years on 8 January 2021.

The authors have previously analysed his tweets on climate change (Allen and McAleer 2018a), on nuclear weapons and Kim Jong-Un (Allen and McAleer 2018b), and have contrasted his first State of the Union Address (SOU) with the previous one by President Obama (Allen et al. 2018). We have also compared some of his speeches with those of Obama and Hitler (Allen et al. 2019a, 2019b).

This paper features an analysis of former President Trump's recent tweets on COVID-19. The tweets are analysed by means of various data mining techniques, including sentiment analysis. The intention is to explore the contents and sentiments of the messages contained, the degree to which they differ, and their potential implications for the national reaction to COVID-19. The data set or corpus includes 159 tweets on the coronavirus that are sourced from the Trump Twitter Archive running from 24 January 2020 to 2 April 2020.

The analysis is performed via the application of a variety of R packages. These include 'tm', a text mining package, created by Feinerer and Hornik (2019), 'Textmining', Eder and Melcer (2016), 'tidytext', by Silge and Robinson (2016), and 'stringi', Gagolewski (2020). We also used 'syuzhet', a sentiment extraction tool, incorporated into an R package by Jockers (2015), 'wordcloud' by Fellows (2018), 'TwitterR', Gentry (2015), and 'tm', Feinerer and Hornik (2019),

Data mining methods are drawn from statistics, machine learning, and database systems and are applied to the analysis of textual data and the exploration of patterns within it. Sentiment analysis features the exploration of the nature of the emotions contained in a text. Differences in sentiment can be viewed in terms of binary distinctions (positive versus negative). Alternatively, different types of emotions can be explored. We used the R packages 'tidytext' and 'syuzhet', which distinguish between eight different emotions, namely trust, anticipation, fear, joy, anger, sadness, disgust, and surprise.

Sentiment analysis has its limitations, who given the mention difficulties interpreting various common language usages, such as sarcasm, negations, and so forth.

It is possible to analyse the sentiments of news feeds using these techniques. Allen et al. (2015); Allen et al. (2017) analyse the influence of the Thomson Reuters News Analytics (TRNA) sentiment series. The first of these papers explored the influence of the Sentiment measure as an asset-pricing factor in pricing DJIA constituent company stocks. The second used an aggregated DJIA market sentiment score and entropy measures, to assess the impact of scores on DJIA returns. Allen et al. (2018) use the Thomson Reuters News Analytics (TRNA) data to successfully augment the Fama–French three-factor model.

Allen and McAleer (2019) undertake an analysis of then President Trump's two State of the Union addresses and also apply Zipf and Mandelbrot's power law to assess the degree to which they differ from common language patterns. In order to provide a contrast and some parallel context, analyses are also undertaken of President Obama's last State of the Union address and Hitler's 1933 Berlin Proclamation.

The structure of these four political addresses is remarkably similar. The three US Presidential speeches are more positive emotionally than is Hitler's relatively shorter address, which is characterized by a prevalence of negative emotions. Hitler's speech deviates most from common speech, but all three appear to target their audiences effectively by use of non-complex speech.

Various papers have explored the influence of news items on asset prices and volatility. Examples of this are given by: Tetlock et al. (2008), Da et al. (2011), Barber and Odean (2008), diBartolomeo and Warrick (2005), Mitra et al. (2009), and Dzielinski et al. (2011). Cahan et al. (2009), and Hafez and Xie (2012), used RavenPack data to examine diversification benefits.

A number of papers provide surveys of this literature. Loughran and McDonald (2014) survey applications of textual analysis and sentiment analysis to the accounting, finance, and economics literature. Kearney and Lui (2014) review sentiment analysis and discuss applications in the related literature.

In this paper, we concentrate on the content of then President Trump's recent tweets on the topic of COVID-19 and their implications for the management of the pandemic, in the early days up to the beginning of April 2020. Does President Trump try to present the pandemic in a positive light for the benefit of his voter base? Does he try to play down the risks of both the virus and the attendant risks? Could the tweets on these topics be interpreted seriously as attempts to avoid scare-mongering?

To provide context, the tweets are presented against a background of what Dr. Fauci was saying in public statements at the time and privately via his recently revealed private emails, as obtained by Buzzfeed on 1 June 2021. It appears, with the benefit of hindsight, that there was an inherent contradiction between his public statements and his private email correspondence. It may be the case that President Trump's instincts were more accurate than was thought at the time, and that his response to the pandemic was not as xenophobic, or 'contrary to the science', as was originally claimed.

(<https://www.buzzfeednews.com/article/nataliebettendorf/fauci-emails-covid-response>). Dr Fauci's dissembling possibly gives the lie to more extreme interpretations of President Trump's pronouncements at the time.

Might President Trump's constant tweeting be fairly described as constituting 'propaganda'? This might be defined as presenting information in a biased or misleading nature, and commonly used to promote a political cause or point of view.

Sentiment analysis will not give a clear answer to this question per-se, but it should reveal patterns in the sentiment displayed within the tweets, correlations and associations in the use of words, and patterns displayed over time in the messaging embedded in the tweets.

Zipf (1932, p. 1) suggested an alternative approach to the analysis of language as a whole based on relative frequency, suggesting: "the accent or degree of conspicuousness of any word, syllable, or sound is inversely proportionate to the relative frequency of that word, syllable, or sound, among its fellow words, syllables, or sounds in the stream of spoken language. As any element's usage becomes more frequent, its form tends to become less accented, or more easily pronounceable, and vice versa".

Zipf (1932) described four important features recognisable in words, 'meaning', 'quality', referring to either positive or negative qualities. The focus of sentiment analysis in this paper. 'Emotional intensity', which can also be related to espousal of sentiment. Plus, what he referred to as being 'order', a concept related to semantic change and the occurrence of the relative frequency of usage of different words.

Zipf suggested a suitable formula for capturing the relative frequency of words is $P_n \sim 1/n^\alpha$, where P_n is the frequency of a word ranked n th, and the exponent α is close to 1. This means that the second most frequently observed word occurs approximately 1/2 as often as the first, and the third word 1/3 as often as the first, and so on.

Mandelbrot (1965) expanded and refined Zipf’s theory concentrating on the view that human languages evolved to optimize the conveyance of information. He drew on Shannon’s (1948) ‘information theory’. Mandelbrot wrote the formula $i(r, k)/k$, where i is defined as the relative number of repetitions of the word $W(r)$ in a sample of length k . This is proposed as being inversely proportional to 10 times r , $i(r, k)/k = 1/10r$.

Shannon (1948, p. 6) showed how artificial languages can be used to approximate natural languages. If all letters are given the same probability and chosen independently, this would be a zero-order approximation. In a first-order approximation, letters would be chosen independently, but their probability of occurrence would match that in the relevant natural language. A trigram structure would be adopted in a third-order approximation. In this case, the probability of each letter would be dependent on the preceding two letters.

Shannon (1948) writes, let $p(B_i)$ be the probability B_i of a sequence of symbols from a source text. Let:

$$G_N = -\frac{1}{N} \sum_i p(B_i) \log p(B_i), \tag{1}$$

where the sum is over all sequences B_i containing N symbols. The implication is that G_N , which is probability mass function, is a monotonically decreasing function of N , and that:

$$\lim_{N \rightarrow \infty} G_N = H.$$

Shannon lets $p(B_i, S_j)$ be the probability of sequence B_i being followed by symbol S_j and $p_{B_i} S_j = p(B_i, S_j)/p(B_i)$ be the conditional probability of S_j after B_i , then let:

$$F_N = - \sum p(B_i, S_j) \log p_{B_i}(S_j), \tag{2}$$

where the summation is over all blocks B_i of $N - 1$ symbols and over all symbols S_j , then F_N is a monotonically decreasing function of N :

$$F_N = NG_N - (N - 1)G_{N-1},$$

$$G_N = \frac{1}{N} \sum_{N=1}^N F_N,$$

$$F_N \leq G_N,$$

and $\lim_{N \rightarrow \infty} F_N = H$.

Shannon (1948) suggests that F_N is the entropy of the N th-order approximation. Mandelbrot (1965) interprets this, derivation of the law of word frequencies, as being consistent with maximising Shannon’s “quantity of information” under certain constraints.

Ficcadenti et al. (2019), represent an application of this type of approach and also review some of the relevant literature. We apply the framework outlined above to analyse President Trump’s tweets on COVID-19 and use them to explore the degree to which the language in them is removed from standard patterns of speech.

The remainder of the paper is divided into four sections. An explanation of the research method is given in Sects. 2, 3 presents the results, and Sect. 4 provides some concluding comments.

Research method

We use a number of R libraries in our data mining and sentiment analysis. These include word cloud, tm, textmining, textreg, and syuzhet, plus a variety of graphics packages. The R package tm provides a basic infrastructure required to organize, transform, and analyze textual data. The process involves importing the body of tweets into a 'corpus'. The corpus, in turn, has to be transformed by various manipulations into a suitable form for analysis. This creates a term-document matrix which can be used for analysis.

When we have the text in matrix form, a large number of R functions (like clustering, classifications, etc) can be used. Associations between words, and their correlations can be examined, and text can be filtered to reveal frequently occurring words. Once we know the frequently occurring words we can create a word cloud, as described by Feinerer and Hornik (2019). Another R library package that is useful for creating and analysing word clouds is 'wordcloud' by Fellows (2018).

For sentiment analysis, we use the R package 'syuzhet' and apply the default syuzet lexicon, which was developed in the Nebraska Literary Lab under the direction of Jockers (2015). The term 'Syuzhet' comes from the Russian Formalists Shklovsky (1917) and Propp (1928), who divided narrative into two components, the 'fabula' and the 'syuzhet'. The first term refers to the 'device' or technique of a narrative, and the second to the chronological order of events. The R package constructs global measures of sentiment into eight constituent emotional categories, namely trust, anticipation, fear, joy, anger, sadness, disgust, and surprise.

To explore how the narrative is constructed, and how the positive or negative sentiment revealed changes over time, we plot the values in a graph, where the x-axis represents the passage of time from the beginning to the end of the text, and the y-axis measures the degrees of positive and negative sentiment.

We develop the appropriate R code to undertake the Zipf and Mandelbrot power law distribution analysis, to assess the degree to which the tweets on COVID-19 deviate from common language, and draw on the R package 'tm'.

The paper features an analysis of then President Trump's public statements via his Twitter account and the application of machine learning techniques. The recent release of Dr. Fauci's emails on the topic gives a more detailed context to President Trump's tweets, both in terms of what Dr Fauci was saying in public, and privately in his email correspondence.

Results and analysis

President trump's tweets

Figure 1 presents a word cloud analysis of President Trump's tweets on COVID-19. It is no surprise that the word cloud suggests that the most frequently occurring word was 'coronavirus' because the tweets in the 'TrumpTwitter Archive' (see: <http://www.trumptwitterarchive.com/archive>) were screened on the word

Table 1 Most frequent word associations

Word	Associated words (correlations)			
Coronavirus	Free (0.36)	Leave (0.32)	Paid (0.32)	Sick (0.32)
China	Closely (0.6)	Agencies (0.48)	Conversation (0.48)	Anywhere (0.48)
China	Good (0.46)	Monitor (0.46)	Ongoing (0.46)	Received (0.46)
China	Top (0.46)	Detail (0.46)	Developed (0.46)	Discussed (0.46)
China	Finished (0.46)	Parts (0.46)	Planet (0.46)	Ravaging (0.46)
China	Understanding (0.46)	Experts (0.38)	Working (0.37)	Virus (0.33)
China	Closed (0.31)	Respect (0.31)	Leading (0.31)	Recovery (0.31)
China	Also (0.31)	Best (0.31)	Developments (0.31)	Large (0.31)
cdc	Including (0.46)	Congratulations (0.40)	Border (0.40)	Closing (0.40)
cdc	Correct (0.40)	However (0.40)	Opposed (0.40)	Soon (0.40)
cdc	Turned (0.40)	Contact (0.40)	Market (0.40)	Relevant (0.40)
cdc	Smart (0.40)	Starting (0.40)	Stock (0.40)	USA (0.40)
cdc	Avoiding (0.40)	cdctravel (0.40)	cdctravelnotice (0.40)	Recommends (0.40)
cdc	Europe (0.36)	Handling (0.35)	Great (0.31)	
President	Vice (0.50)	Airline (0.50)	Ceos (0.50)	Corona (0.50)
President	Impact (0.50)	Met (0.50)	Earlier (0.39)	Discuss (0.37)
President	Mikepence (0.33)			
Busy	Battle (1.00)	Calling (1.00)	Flight (1.00)	Republican (1.00)
Busy	Wasting (1.00)	Ahead (0.81)	Anything (0.81)	Bad (0.81)
Busy	Closings (0.81)	Hoax (0.81)	Putting (0.81)	Wrong (0.81)
Busy	Immigration (0.71)	Border (0.71)	Impeachment (0.71)	Dems (0.70)
Busy	Party (0.63)	Time (0.63)	Way (0.63)	Called (0.53)
Busy	Nothing (0.53)	Else (0.50)	Democrats (0.49)	Look (0.49)
Busy	Make (0.49)	Border (0.40)	Early (0.40)	
Will	Directed (0.32)	Direction (0.32)	Evening (0.32)	Capitalliquidity (0.32)
Will	Chadpergram (0.32)	Firms (0.32)	Loans (0.32)	sba (0.32)
Democrats	Battle (0.49)	Busy (0.49)	Calling (0.49)	Flight (0.49)
Democrats	Republican (0.49)	Wasting (0.49)	Party (0.46)	Ahead (0.40)
Democrats	Anything (0.40)	Bad (0.40)	Border (0.40)	Closings (0.40)
Democrats	Hoax (0.40)	Putting (0.40)	Wrong (0.40)	Nothing (0.38)
Democrats	Immigration (0.35)	Blamed (0.35)	Ended (0.35)	Fault (0.35)
Democrats	Slowly (0.35)	Trump (0.35)	Border (0.35)	Impeachment (0.35)
Democrats	Pulled (0.35)	Really (0.35)	Republicans (0.35)	See (0.35)
Democrats	Teamwork (0.35)	Needs (0.35)	Someone (0.35)	Blaming (0.35)
Democrats	Scam (0.35)	Talking (0.35)	Charliekirk (0.35)	Incite (0.35)
Democrats	Let (0.35)	Mainstream (0.35)	Straight (0.35)	Trying (0.35)
Democrats	Comcast (0.35)	Covers (0.35)	Disinformation (0.35)	Harm (0.35)
Democrats	Horribly (0.35)	Looking (0.35)	Put (0.35)	Dems (0.34)

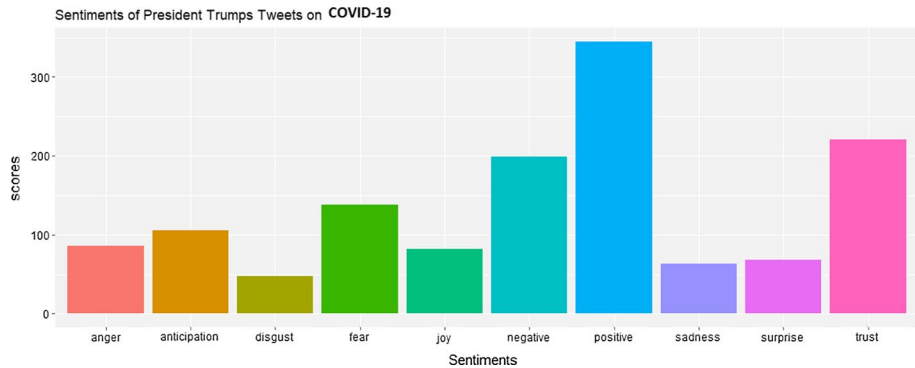


Fig. 3 Sentiment analysis of president trump's tweets on the COVID-19

economic effects of shutdowns and enforced social distancing. Words with high correlations include 'free', 'leave', 'paid', and 'sick'.

The correlations with 'china' are surprisingly positive or neutral, and include 'closely', 'agencies', 'conversation', 'anywhere', 'good', 'monitor', 'ongoing', 'received', 'top', 'detail', 'developed', 'discussed', 'experts', 'respect', 'best', 'leading', to mention a few. The only possibly perjorative terms appear to be 'ravaging' and 'virus'.

The word 'president' is associated also with the 'vice' president, Mike Pence, and words such as 'airlines', 'ceos', 'corona', 'impact', 'met', and so on. Another frequently occurring word in this set of tweets is 'busy', which seems to be highly associated with 'battle', 'calling', 'flight', 'republican', 'wasting', 'ahead', 'anything', 'bad', 'closing', 'hoax', 'putting', 'wrong', 'immigration', 'border', 'impeachment', 'dems', just to mention just those words with correlations in excess of 0.70.

With respect to the previously mentioned 'democrats', there seems to be a much greater prevalence of negative associations than those mentioned with respect to the previous words considered. Some of the words with negative connotations include 'battle', 'wasting', 'bad', 'closings', 'hoax', 'wrong', 'nothing', 'blamed', 'fault', 'impeachment', 'scam', 'incite', 'harm', 'disinformation', and so forth. The use of President Trump's twitter feed as a political weapon seems to be apparent in relation to the word 'democrats'.

The sentiment analysis of these tweets shown in Fig. 3 reveals that overall they are predominantly positive, as revealed by the two central columns in the figure. The predominant emotion conveyed is trust, followed by the negative emotion fear. These are then followed by anticipation, joy, anger, surprise, sadness and disgust, in order of their relative predominance.

Figure 4 plots the 'emotional valence' of this series of tweets. This refers to the pattern of sequential positive and negative emotions displayed as the tweets on corona virus unfold through time. The plot of these patterns shown in Fig. 4 does not reveal a particular pattern in the occurrence of positive and negative emotions.

Figure 5 shows a theoretical application of Zipf's law to the set of tweets. A full confirmation of Zipf's law would show a line of slope of negative one in the plot in Fig. 5, running from the top left to the bottom right. In the diagram, the y axis depicts the logarithm of relative frequency and x axis the logarithm of the index.

The regression model which produced the diagram in Fig. 5 is shown in Table 2. The plot deviates from a theoretical plot of a line with a slope of negative 1. A flatter Zipf slope

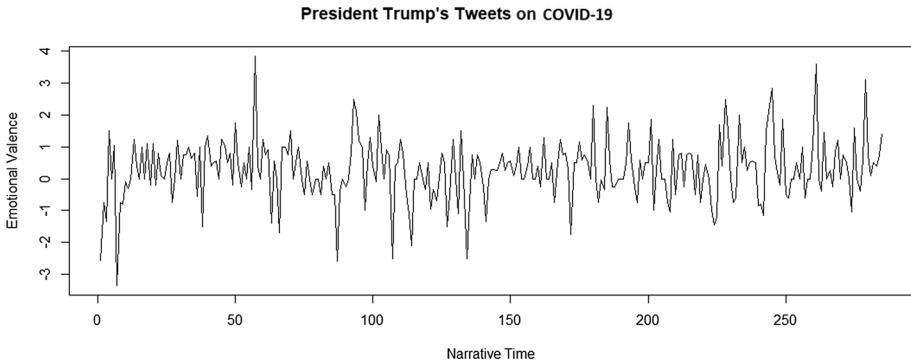


Fig. 4 Emotional valence of tweets on COVID-19

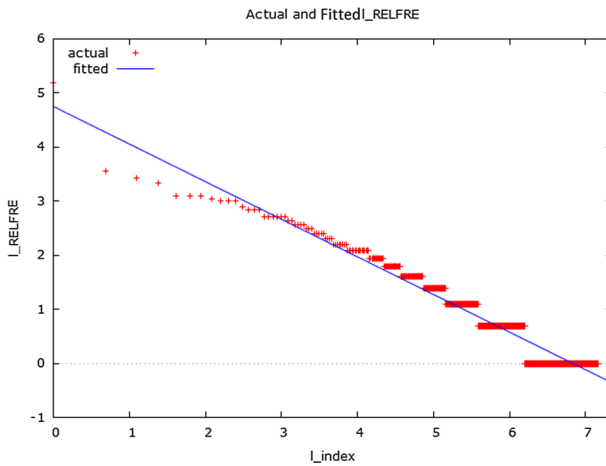


Fig. 5 Estimation of Zipf law relationship

Table 2 Zipf Regression Model 3: OLS, using observations 1–1296 Dependent variable: I_RELFRE

	Coefficient	Std. Error	<i>t</i> -ratio	<i>p</i> -value
[lex] const	4.75172	0.0315005	150.8	0.0000
I_index	− 0.695792	0.00504085	− 138.0	0.0000
Mean dependent var		0.458327	S.D. dependent var	0.710410
Sum squared resid		41.56553	S.E. of regression	0.179225
<i>R</i> ²		0.936402	Adjusted <i>R</i> ²	0.936353
<i>F</i> (1, 1294)		19052.45	<i>p</i> -value(<i>F</i>)	0.000000
Log-likelihood		390.0244	Akaike criterion	− 776.0488
Schwarz criterion		− 765.7148	Hannan–Quinn	− 772.1709

can indicate a more random signal, but it can also indicate a broader vocabulary that conveys a more precisely worded message. Zipf suggests that attempts to remove ambiguities should produce a flatter slope that favours the recipient. The estimated slope coefficient is -0.70 , which is highly significant at the 1% level, with a t -statistic of -138 and an adjusted- R squared value of 0.94 .

The slope estimate suggests that President Trump's tweets on COVID-19 are designed to favour the recipient, and deviates from standard language patterns. This makes perfect sense, given the sparse nature of tweets and the fact that they are frequently designed to convey a simple message.

Dr Fauci's public statements and private emails

In an email response to Danelle Steinberg on 3 February 2020, Dr Fauci replied: "You ask that there have been animal markets for a long time, and so why now. The fact is that this is likely pure chance +/- more interactions in the human-animal interface. Animal viruses mutate and most of the time the mutations have no significant impact on virus transmission to humans. Sometimes they mutate and allow single "dead end" transmissions to individual humans with no efficiency in going human to human and so we get individual infections and no outbreak as we have seen with H5N1 and H7N9 influenzas that jump from chickens to humans but do not go from human to human. Then rarely, animal viruses mutate and the mutation allows them not only to jump species to humans, but to also efficiently spread from human to human". Dr Fauci appeared to be very reluctant to consider that the virus may have escaped from a laboratory in Wuhan.

Dr. Fauci has also been a proponent of the continuation of gain of function research and in a piece written in (2012), stated that: "Scientists working in this field might say—as indeed I have said—that the benefits of such experiments and the resulting knowledge outweigh the risks. It is more likely that a pandemic would occur in nature, and the need to stay ahead of such a threat is a primary reason for performing an experiment that might appear to be risky."

Buzzfeed notes that Dr. Fauci responded to Sivia Burwell, who had emailed him a query on 5 February 2020 about the efficacy of wearing a mask, with the comment: "Masks are really for infected people to prevent them from spreading infection to people who are not infected rather than protecting uninfected people from acquiring infection." It does appear that Dr. Fauci's advice about masks did change, with his endorsement of the wearing of one or two masks increasing over time.

The NIH has come under significant criticism in recent weeks over funding WIV research relating to change-of-function, while Dr Fauci has denied that the National Institute of Health (NIH) has funded gain of function research at the Wuhan Institute of Virology (WIV). He told a US Senate hearing that the NIH "has not ever and does not now fund gain-of-function research in the WIV". Yet researchers at WIV, including its prominent virologist Dr. Shi Zhengli have disclosed that work on coronaviruses had been funded by NIH grants.

It does appear that Dr. Fauci had a vested interest in steering attention towards the hypothesis that COVID-19 jumped from animals to humans in the Wuhan wet market, given that the NIH had actually been funding gain of function research via Dr. Peter Daszak and the EcoHealth Alliance, a US-based organization that conducts research and outreach programs on global health, conservation and international development.

Dr Daszak complained in *Nature* on 21 August 2020 about 'unfounded rumours' that the COVID-19 pandemic was caused by a coronavirus released from its laboratory. The NIH cancelled EcoHealth Alliance's grant in April 2020, days after US President Donald Trump told a reporter that the United States would stop funding work at the WIV. The actual origins of COVID-19 remain uncertain, but the lab leak hypothesis has been gaining traction at the time of writing.

Conclusion

In this paper we have analysed a set of 159 tweets on COVID-19 that are sourced from the Trump Twitter Archive, running from 24 January 2020 to 2 April 2020. We have used a variety of R library packages to analyse the tweets using different data and text mining routines, with a focus on sentiment analysis and applications of the Zipf law.

The analysis reveals that former President Trump used the Twitter feed to deliver effectively a simplified set of repeated messages. Given the nature of the topic, it is perhaps surprising that sentiment analysis reveals that the general tenor of the tweets is largely positive. This confirms previous findings in analyses of climate change Allen and McAleer (2018a), and on nuclear weapons and Kim Jong-Un Allen and McAleer (2018b). A positive note was also evident in the contrasts we undertook of then President Trump's first State of the Union Address (SOU) with the previous one by President Obama (Allen et al. 2018), together with some of his speeches with those of Obama and Hitler (2019).

The current analysis reveals that former President Trump's tweets do not appear to be highly critical of China, but show strong evidence of a political theme, and are highly critical of the Democrats. The Zipf analysis suggests that the tweets contain simplified language to convey the simple and effective political message to the target audience. Then President Trump continued to make effective use of Twitter to target his political enemies and deliver questionable information, while avoiding the filters provided by conventional TV and news media. This avenue of effective communication ceased when he was banned for at least two years from Twitter and Facebook in early 2021.

The recent controversy about the nature of Dr. Fauci's private emails and the purported contrast between his private and public statements suggest that then President Trump's criticism of Dr Fauci may have had more justification than was realized at the time, and his decision to stop gain of function research funding may have been a wise move.

Acknowledgements The authors are particularly grateful to Professor Wolfgang Glänzel, Editor in Chief, and to anonymous reviewers for helpful comments and instructions re the paper. We note that the method of analysis used in the paper is the same as in Allen et al. (2019a), and the accompanying working paper, Allen et al. (2016b).

Funding Michael McAleer is most grateful to the Australian Research Council, Ministry of Science and Technology (MOST), Taiwan, and the Japan Society for the Promotion of Science, Grant Number MOST 108-2410-H-005 -008 for financial support.

Declarations

Ethical approval This article does not contain any studies with human participants performed by any of the authors.

References

- Allen, D. E., McAleer, M., Singh A. K. (2015). Machine news and volatility: The dow jones industrial average and the TRNA real-time high frequency sentiment series, chapter 19. In G. N. Gregoriou *Handbook of high frequency trading*. Elsevier, Academic Press.
- Allen, D. E., McAleer, M., & Singh, A. K. (2017). An entropy-based analysis of the relationship between the DOW JONES index and the TRNA sentiment series. *Applied Economics*, 49, 677–692.
- Allen, D. E., & McAleer, M. (2018). Fake news and indifference to scientific fact: President trump's confused tweets on global warming. *Climate Change and Weather, Scientometrics*, 117(1), 625–629.
- Allen, D. E., & McAleer, M. (2018). President Trump tweets supreme leader Kim Jong-Un on nuclear weapons: A comparison with climate change. *Sustainability*, 10(7,2310), 1–6.
- Allen, D. E., & McAleer, M. (2019a) Fake news and propaganda: Trump's Democratic America and Hitler's National Socialist (Nazi) Germany. *Sustainability*, 11(19), 5181.
- Allen, D. E., McAleer, M. J. (2019b). Fake news and propaganda: Trump's Democratic America and Hitler's National Socialist (Nazi) Germany (No. EI2019-17). Econometric Institute Research Papers. Retrieved from <http://hdl.handle.net/1765/115615>
- Allen, D. E., McAleer, M., Reid, D. M. (2018). Fake news and indifference to truth: Dissecting tweets and state of the union addresses by presidents Obama and Trump. *Advances in Decision Sciences*, 22(A), 1–23.
- Barber, B. M., & Odean, T. (2008). All that glitters: The effect of attention and news on the buying behaviour of individual and institutional investors. *Review of Financial Studies*, 21(2), 785–818.
- Cahan, R., Jussa, J., Luo Y.(2009). Breaking news: How to use news sentiment to pick stocks, MacQuarie US Research Report.
- Da, Z. H. I., Engelberg, J., & Gao, P. (2011). In search of attention. *Journal of Finance*, 66(5), 1461–1499.
- diBartolomeo, D., Warrick S. (2005). Making covariance based portfolio risk models sensitive to the rate at which markets react to new information. In J. Knight and S. Satchell. (Eds.), *Linear factor models*, Elsevier Finance.
- Dzieliński, M., Rieger, M. O., & Talpsepp, T. (2011). volatility asymmetry, news, and private investors. *Handbook of news analytics in finance* (pp. 255–270). Wiley.
- Eder, M., Melcer, T. (2016). Textmining: Integration of text mining and topic modeling packages. R package version 0.0.1. <https://CRAN.R-project.org/package=textmining>
- Gagolewski, M. et al. (2020). R package stringi: Character string processing facilities. <http://www.gagolewski.com/software/stringi/>
- Gentry, J. (2015). twitterR: R based twitter client. R package version 1.1.9. <https://CRAN.R-project.org/package=twitterR>
- Fauci, A. S. (2012). Research on highly pathogenic H5N1 influenza virus: The way forward. *American Society for Microbiology*. <https://doi.org/10.1128/mBio.00359-12>
- Feinerer, I. Hornik K. (2019), tm: Text mining package, R package version 0.7-7. <https://CRAN.R-project.org/package=tm>
- Fellows, I. (2018), wordcloud, <https://CRAN.R-project.org/package=wordcloud>
- Ficcadenti, V., Cerqueti, R., & Ausloos, M. (2019). A joint text mining-rank size investigation of the rhetorical structures of the US presidents' speeches. *Expert Systems with Applications*, 123, 127–142.
- Hafez, P., Xie, J. (2012). Factoring sentiment risk into quant models, RavenPack International S.L.
- Jockers, M. L. (2015). Syuzhet: Extract sentiment and plot arcs from text. <https://github.com/mjockers/syuzhet>.
- Kearney, C., & Lui, S. (2014). Textual sentiment in finance: A survey of methods and models. *International Review of Financial Analysis*, 33, 171–185.
- Loughran, T., & McDonald, B. (2016). Textual analysis in accounting and finance: A survey. *Journal of Accounting Research*, 54(4), 1187–1230.
- Mandelbrot, B. (1965). Information theory and psycholinguistics, Chapter in. In B. B. Wolman (Ed.) *Scientific psychology*. Basic Books, New York.
- Mitra, L., Mitra, G., & diBartolomeo, D. (2009). Equity portfolio risk (Volatility) estimation using market information and sentiment. *Quantitative Finance*, 9(8), 887–895.
- Propp, V. (1928). *Morphology of the folk tale*, English trans. Laurence Scott. TX: University of Texas Press (first published in Moscow in 1928).
- Shannon, C. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 379–423, 623–656.
- Shklovsky, V. (1917), Art as Technique. In L. T. Lemon, M. Reis (Eds.) (1965) *Russian formalist criticism*, University of Nebraska Press.

- Silge, J, Robinson, D. (2016). Tidytext: Text mining and analysis using tidy data principles in R, *_JOSS_*, *1*(3). <https://doi.org/10.21105/joss.00037>
- Tetlock, P. C., Macskassy, S. A., & Saar-Tsechansky, M. (2008). More than words: quantifying language to measure firms' fundamentals. *Journal of Finance*, 63, 1427–1467.
- Zipf, G. K. (1932). *Selected studies of the principle of relative frequency in language*. Cambridge, Mass: Harvard University Press.