Comparison of Account Survival on Twitter During the First Wave of COVID-19 in Four Countries

Balint Gyevnar

MInf Project (Part 2) Report
Master of Informatics
School of Informatics
University of Edinburgh

2021

Abstract

The user accounts on online social networks (OSNs) are constantly changing. New ones are created at an increasing pace, while others are deleted by users, protected from public view, or suspended by the online service providers.

The study of OSN account survival can shed light on aspects such as OSN adoption and abandonment, incidence of bots, or enforcement of OSN platform rules, but the ongoing pandemic has significantly altered the online behaviour of people. Therefore, in this thesis I compare the extent to which the first wave of the COVID-19 pandemic between March and April 2020 affected account survival patterns in four countries with different approaches to managing COVID-19: the United States, the United Kingdom, Singapore, and New Zealand.

The data set for the analysis was obtained from the Twitter Streaming API between March 10 and April 29, 2020. It consists of 17.3M unique tweets in English by 15.3M unique accounts. The thesis presents a detailed quantitative analysis of this data using statistical methods.

The results reveal significant and consistent differences among patterns of account survival. Overall trends are similar across the four countries studied, with suspended accounts being the most active. Surprisingly, the frequency of COVID-19 mentions was largely unaffected by case-counts, and mostly followed a steady downward trend over time. For the US and the UK, COVID-19 mentioning patterns were similar, regardless of account type. However, New Zealand showed substantial variation over time, while suspended accounts from Singapore were less likely to mention COVID-19 than all other account types.

Based on the results of the thesis, further temporal and qualitative studies of account survival can be undertaken, while additional comparisons across various socio-economic variables could shed light on digital exclusion and bias.

The thesis concludes that account survival type should be taken into account when examining how the pandemic is discussed on Twitter.

Acknowledgements

I would like to thank my supervisor Maria Wolters for guiding me through the ardous task of writing this thesis and making it into a paper as well.

Table of Contents

Chapter 1

Introduction

1.1 Account Survival

Online social networks (OSNs) form an important part of modern society and their usage increased significantly during the COVID-19 pandemic[1]. The abundance and quality of OSN data has for a long time enabled researchers to analyse a very wide range of social and psychological questions as well [32].

In OSN ecosystems user accounts have particular significance because they serve as the basic method of access to most platforms. However, one largely understudied aspect of user accounts is how their visibility and accessibility changes over time. That is, how their **survival state** evolves.

Accounts on OSNs do not persist across time [57]: new ones are created at an increasing pace, while others may get deleted by users, or protected from public access. Accounts can also be suspended by the online service providers most often due to spamming or security breaches. This thesis builds on the previous work of Volkova and Bell [57], who outlined three main account types depending on survival state:

1. **Alive**: Alive accounts, that remain publicly accessible by any other registered user and form the widest, baseline group.
2. **Deleted**: Inaccessible accounts removed by their owners themselves.
3. **Suspended**: Inaccessible accounts suspended by service providers, usually for spamming or security reasons.

Though this categorisation covers the majority of accounts on OSNs, it fails to account for users who wish to remain a member of the social network but want to hide their activity in some way. Therefore, we extend this classification with a fourth category, **protected**. Protected accounts are still active, but their content can no longer be accessed by the public. Finally, we refer collectively to the constant process of accounts changing survival types on OSNs as **account survival (AS)**.

[1] https://www.nytimes.com/interactive/2020/04/07/technology/coronavirus-internet-use.html

1.2 Motivation

1.2.1 Twitter & COVID-19

In this thesis, Twitter was used to collect data. Twitter is a micro-blogging service and one of the largest OSNs by daily traffic [27]. It provides users a fast, real-time and personalised way to express their thoughts or to follow others through short 280-character-long messages called *tweets*.

Tweets give a detailed snapshot into people's thought processes or opinions and data from Twitter is used extensively to answer a wide range of research questions in several fields [64]. It also presents an excellent opportunity to study account survival [57].

Furthermore, several studies have already showed a shift in OSN behaviour attributed to the COVID-19 pandemic. Mentions of anxiety and depression increased [19], ageism became more prevalent [41], and the wide spread of fake news and misinformation campaigns culminated in an ”infodemic” [13].

Twitter is being used extensively in this context to express reactions to the ongoing events and changes. While many studies have already used data from Twitter to capture important information about the effects and spread of COVID-19 (e.g. [49, 1, 36]), however account survival has been largely neglected in the literature so far.

Finally, there is large variability in the socio-economic effects of the pandemic across spatial dimensions as well [42, 52]. Different countries have adopted different emergency measures which has an impact not only on case and death counts, but can also be reflected in people's behaviour, which in turn may have profound effects on their OSN usage as well. Therefore, it is also beneficial to establish to what extent do these differences affect account survival in various regions.

1.2.2 Privacy

Most websites offer the possibility to opt-out of their services, to remove personal data, or to completely delete an account, however these functions are often made deliberately difficult to find [22] as it may be financially detrimental to the service providers [30]. This leads to skepticism and doubt about the trustworthiness of these companies [22].

With revelations around private data mismanagement in the ”big-tech” industry (e.g. Cambridge-Analytica scandal) and people's increasing desire to control their online presence [23], it is becoming ever more important to understand what actually drives people to delete or protect their accounts on OSNs.

However, a gap in explaining how users are motivated to hide their online presence was highlighted in an empirical study by Hinds et al. [24], where they showed that very few users actually changed their online behaviour in reaction to the Cambridge-Analytica scandal. This phenomenon is known as the *privacy paradox* [8] and studying account survival could help explain its underlying causes.

Additionally, in the wake of COVID-19, countries decided to use a mixture of documentation, modelling and contact tracing solutions for the prevention of further spread-ing of the virus, however these practises have raised concerns about handling of private data [45, 47].

The regional differences are significant in COVID-19 contact tracing as well. For example, Singapore has effectively made it compulsory for its citizens to use its tracking app ”TraceTogether” thereby creating an efficient but potentially abusable tracking solution, while the UK’s ”Test and Trace” approach was shown to be largely underutilised [44].

1.2.3 Online Mental Health

There is a large body of literature studying the effects of technology use and disuse in the field of *technostress* [5]. For example, evidence suggests that social media *discontinuance* is closely tied with information and social *overload* [18]. While most research in this direction is focused on applying social and psychological theories to explain technology discontinuance, directly measuring account survival could provide concrete evidence to validate these studies and to point towards other directions for future research.

In the context of the COVID-19 pandemic, account survival is of particular interest. Social media use may exacerbate existing mental health problems [25], and therefore, deleting one’s social media account can be a positive step towards preserving mental well-being.

Studies showed that COVID-19 had a direct effect on the self-disclosure of users [43] while the global misinformation about the pandemic (”infodemic”) has led to social media fatigue [28].

1.2.4 Spamming & Automated Accounts

While previous motivations for this thesis were mostly focused on account deletion or protection, account suspension relates to different underlying factors as suspension is done by the OSN itself. Twitter usually suspends accounts due to three main reasons:

1. They violated Twitter’s policy on platform manipulation and spam[2]
2. It is a safety measure as the accounts may be at risk (e.g. due to data breach)
3. The accounts engaged in violent or abusive behaviour.

There is a large literature on understanding, detecting and classifying suspended accounts on Twitter as they have close ties to spam and automated accounts (e.g. [39, 14]). During the ”infodemic” of COVID-19 suspended accounts may have exerted significant influence to societies through systematic misinformation campaigns, which prompted an increase in research interest for the analyses of spamming and bots [15, 56].

[2] https://help.twitter.com/en/rules-and-policies/platform-manipulation

1.2.5 Research Methodology

OSN data is not persistent so an important benefit of understanding account survival lies in the ability to create reproducible datasets through the exclusion of to-be-altered users [57]. This could enable researchers to share datasets that are longer-lasting and publish results that are more easily reproducible.

1.3 Problem Statement

This work presents an analysis of account survival during the first wave of the pandemic using a COVID-19 specific dataset scraped from Twitter. Since the responses to the pandemic varied considerably across the world, we sampled data from four countries with different COVID-19 responses: the United States of America, United Kingdom, Singapore, and New Zealand. The goal of the thesis is to answer the following research question:

- **RQ1:** What are the main characteristics of different account survival types during the observed time period?
- **RQ2:** How do the mentions of COVID-19 evolve over time across various survival types?
- **RQ3:** Do changes in the prominence of COVID-19, such as infection rate, affect the engagement with COVID-19 related across survival types?

The results showed significant differences among AS types, while case-counts had next to no effects on the interest in COVID-19. The addition of protected accounts highlighted the existence of a qualitatively different user-group that has received little attention so far from the scientific community. It can be concluded that account survival is an important factor to consider when studying the discussion of COVID-19 on Twitter.

This thesis formed the basis of a paper submission to the ACM Conference on Web Science in 2021 (WebSci'2021). At the time of writing, the review process was still ongoing.

1.4 Previous Work

This thesis is a continuation of the initial investigation [21] on patterns of account deletion and suspension (i.e. account survival) on online social networks (OSNs), in particular using Twitter data. We used usernames as the basis for our analysis, as they provide a persistent and unique identifier that were shown to encode rich semantic information. Account survival is a neglected topic among contemporary social network research, however it was shown that useful information can be extracted from survivability that help with the detection of automated or spam accounts.

We first performed an exploratory analysis on usernames that gave several insights into their syntax and semantics. We then presented and compared four methods for extract-ing knowledge about accounts from usernames. This was done through segmentation, and the methods included n-grams, random splitting, regular expressions incorporating domain knowledge, and unsupervised morphological segmentation.

For predicting the future survival of an account, we set up a bag-of-words (BoW) feature space based on the segmented usernames. Due to the size and sparsity of this space a 2-fold dimensionality reduction was performed. We also considered additional features based on metadata extracted from user accounts and tweets.

Finally, we used two standard machine learning algorithms, Support Vector Machines (SVMs) and Random Forests (RFs) given our features as input and survivability as label. We found that the dimensionality reduced BoW model of usernames was able to improve on majority-predicting baseline performance, however it was significantly outperformed by other models that also incorporated metadata.

We concluded, that usernames show potential to be used for survival prediction, however their use requires large datasets or better representational approaches to alleviate sparsity.

In this work, we continue our investigation of account survival on Twitter, however we specifically focus on accounts that have engaged with the COVID-19 pandemic, and we base our analysis on both account and tweet data, instead of focusing on usernames.

Chapter 2

Background

2.1 Twitter

The following is a short description of core Twitter concepts. Twitter's main strength comes from the ability for users to form relations with other users, which are called *friend-follower* relations: users may subscribe to the Twitter feed of another account. This account then becomes a *friend* of the user, and the user becomes a *follower* of the account.

Users can also directly mention others through *tagging* with the @ symbol, which is also how Twitter represents *comments* to tweets. Users are able to verbatim repost someone's tweet on their own feed through *retweeting*. Topics or ideas in tweets can be marked with *hashtags* using the # symbol. Furthermore, a user can group accounts into *lists* based on any criteria.

Tweets can contain any mixture of text, links, media, or polls. The setup of this ecosystem generates syntactically and semantically rich data while also representing network and interaction information. This data can be simply accessed using Twitter's publicly available application programming interface (API).

Topics investigated using Twitter data include sentiment analysis [35], depression [55] and stance detection [4], analysis of trolls [37] and conspiracy theories [33] or real-world event detection [48].

2.2 Account Survival

Account survival is a largely understudied aspect of OSNs, but most similar to our study is the work of Volkova and Bell [57]. They gave a detailed description of effective signals to predict the future survival state of an account with shallow, linguistic, and network features. They also compared their results across three languages and showed that languages have to be taken into consideration when performing cross-country analyses of account survival.

2.2.1 Suspended Accounts

Most of the attention for account survival has been directed towards suspended accounts in the literature. For example, Chowdhury et al. [11] carried out a large retrospective analysis of suspended accounts and found that a large number of them went undetected for a long time. During this period, they managed to build extensive relationship networks and formed distinct communities. They also discovered dormant groups of suspended accounts that were created en masse possibly with the intention to manipulate the platform.

Additionally, Wei et al. [59] showed that suspended accounts constitute an essential element of mention and co-topic network-topologies on Twitter. These topologies can turn brittle as we remove suspended accounts suggesting that these accounts have a profound effect on what users perceive to be influential. They also distinguished between different types of suspended accounts, such as spam bots, social bots, and human extremists.

Suspended accounts have also been studied on detecting online spammers and bots. For instance, Chu et al. [12] were one of the first who outlined a system using shallow features and entropy measures to predict whether an account is controlled by a human, a computer, or both. They observed that bot accounts were more likely to be suspended than human accounts.

Furthermore, Cresci et al. [14] gave a classification of bot accounts based on their behaviour. They showed a paradigm-shift of automated accounts, and highlighted a new kind of bot which tries to mimic human behaviour. They called these accounts social bots. They pointed out, that the appearance of social bots are likely a response to the fact, that traditional bot accounts are more easily suspended.

2.2.2 Account & Tweet Deletion

In contrast to suspension, account deletion and account protection have not been studied in detail, though content removal from OSNs has received some attention. In a study on tweet removal, Almuhimedi et al. [2] identified several dimensions where deleted and 'un-deleted' tweets differed significantly. These included the client used, location of posting, and number of replies received. They also hypothesised, that location privacy and linguistic correctness are among the main reasons that motivate tweet deletion, however concluded that more investigation was necessary.

In a relevant study focusing on deleted accounts, Zhou et al. [63] found three groups of accounts through clustering based on deletion tendencies: normal users, bulk-deletion users, and 'hyperactive' users with large volumes of posts. Using this grouping, they identified a set of five features that showed good performance when identifying 'regrettable' deleted tweets.

Bastos [9] showed that during the Brexit debate in 2016 one in three low-quality politically-charged posts were deleted compared to the overall tweet decay rate of 4% on Twitter. They also suggested, that controversial or political posts are much more likely to be deleted in general. A larger deletion rate during the Brexit vote was also independently confirmed by Llewellyn et al. [37], who also pointed out that much of the deleted activity came from trolls and automated accounts.

2.2.3 Protected Accounts

The study of protected accounts on Twitter has received very little attention so far. Relevant studies on user account protection are usually concerned with privacy preservation and designing systems resilient to privacy threats [6, 16].

Zheleva and Getoor [61] showed, that the privacy features of OSNs are often not enough to effectively protect users. In proving their hypothesis, they created a system based on the idea of homophily and network measures. They were able to successfully and effectively reconstruct sensitive and protected information of users.

Kekülüoğlu et al. [34] showed that protected accounts engaged more actively with posts mentioning significant life-events. In particular, they highlighted that shadow accounts and replies can be easily used to reconstruct the topic of discussion of protected tweets, which also showed the insufficiency of account protection of OSNs

Additionally, Waniek et al. [58] reinforced the difficulty of protecting the privacy of OSN accounts by studying the computational complexity of disguising oneself on OSNs. They showed that the task of actively avoiding OSN analysis tools by altering relationship networks is NP-hard according to several measures of node centrality. Though they proposed heuristics to overcome this issue, they also pointed out that discriminated groups are much harder to mask.

2.3 COVID-19 Response of Selected Countries

We examine the effects of the pandemic along spatial dimensions by performing a cross-country comparison of our results. The following section provides an overview of the course of the pandemic in each of the selected countries and the different approaches enacted to combat COVID-19. Our data was obtained during the first wave of the pandemic, so this section is constrained to that time period. Our data for case and death counts comes from the European Centre for Disease Prevention and Control[1].

We selected four English-speaking countries based on a combination of several factors that cover a diverse range of COVID-19 responses: the United States of America, United Kingdom, Singapore, and New Zealand. The factors considered during selection were as follows:

1. Efficacy, extent, and enforcement of preventive measures
2. Strictness and enforcement of border control
3. Cultural factors and compliance with measures
4. Case and death counts

[1]https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographic-distribution-covid-19-cases-worldwide

2.3.1 United States of America

The pandemic response in the USA was repeatedly hit with hurdles on a federal level. National emergency was only declared on March 13 and in-country travel restrictions were introduced only later as well. Large differences emerged on a state-level, with some states opting for full lockdowns and introducing preventative measures, while others maintained pre-pandemic conditions. For example, California issued stay-at-home notices on 19 March, while Alabama followed only on April 4. Early easing of restrictive measures began around the end of April across the country. Views on the necessity of non-pharmaceutical measures became divided across cultural, ethnic and political lines. Combined with governmental miscommunication, this disorganised response resulted in the fastest growing case counts and mortality rates globally [7, 31].

2.3.2 United Kingdom

After an initial attempt at achieving ”herd immunity” was quickly abandoned, the government introduced strict lockdown measures on March 23. This managed to break the rapid spread of the pandemic, with the population largely adhering to regulations. Gradual easing of measures began on May 10 as the government outlined detailed plans for lifting the lockdown. However, the initial delay in response, a continuous governmental miscommunication about the pandemic, and a lack of effective travel restrictions resulted in one of the largest death rates and case counts globally [26, 3].

2.3.3 Singapore

Deriving from its previous experience with the SARS outbreak in 2003, Singapore swiftly activated a stringent 4-tier system of pandemic spread prevention (DORSCON) with level-3 restrictions entering into force on February 7. The measures included strictly enforced lockdowns, mandatory mask-wearing, and aggressive testing practices. On March 20, Singapore was also the first in the world to introduce a mobile contact-tracing solution TraceTogether. The country experienced a quick rise of case counts in April due to infections within foreign workers' dormitories, however the preventive measures, a complete border lockdown, and clear communication from the government stopped the spread of COVID-19 and kept the infection counts low afterwards. These restrictions were eased starting from May 2 [60, 51].

2.3.4 New Zealand

Early implementation of the highest COVID-19 alert-level by March 23, combined with a total travel ban enacted on March 19, and the compliance of the population with preventative measures, such as mask wearing and social distancing, lead to a quick drop in case counts and low mortality rates. This also resulted in the complete elimination of community transmissions by June 8. The fast and effective response lead to a quick easing of protective measures by April 27, which minimised the negative mental and economic effects [29, 17].

Chapter 3

Methods

3.1 Data Collection

Scraping of data began as soon as the outline for the thesis was formulated. It started on March 10 and ran until April 29 which constitutes a 50-day time-period. Data collection was stopped when interest in the pandemic on Twitter has visibly subsided, that is, no new COVID-19-related phrases appeared among the top-ten trending topics.

The dataset was scraped using the Streaming API of Twitter[1] accessed through a Python application that used the Twython package[2]. The Streaming API provides a continuous sample of 1% of all tweets that are published on the platform. The raw feed was filtered using a list of COVID-19-related keywords. This list was then continuously expanded every day by including new trending words or hashtags that are related to COVID-19. All retrieved tweets were stored from this filtered stream along with their issuing account's information. An excerpt from the final list of keywords is given in Table 3.1. Finally, the retrieved data was stored in an encrypted relational database based on PostgreSQL[3]. The fields of the tables match the keys of the objects returned by the Twitter API[4].

The following are a few summary statistics about the scraped dataset. In total, about $\sim$ 15.3M unique accounts and $\sim$ 58.3M tweets were collected of which $\sim$ 38.9M tweets were retweets. On average a user has 5.32 ± 23.04 (Min: 1; 25%: 1; Median: 2; 75%: 3; Max: 7,313) tweets stored in the dataset. The data collection process collected tweets primarily in English ($\sim$ 96% of all tweets), however there are a total of 65 languages in the dataset. Among the chosen countries of Section 2.3, most of the accounts were based in the USA, while New Zealand had the least occurrences.

[1] https://developer.twitter.com/en/docs/twitter-api/v1/tweets/filter-realtime/overview
[2] https://twython.readthedocs.io/en/latest/index.html
[3] https://www.postgresql.org/
[4] https://developer.twitter.com/en/docs/twitter-api/v1/data-dictionary/overview/{user,tweet}-object

coronavirus	coronaviruschallenge	coronaviruschina
stayhome	coronacrisis	coronavid19
corona	stayhomesavelives	COVID2019
COVID19	staysafestayhome	stayhomechallenge
COVID	COVID19Walkout	covidiot
theworldendedandgodsaid	herdimmunity	handsanitizer
highriskcovid19	socialdistancing	myquarantineinsixwords
herd_immunity	isolationessentials	viruscorona
lockdown	ncov	panicbuying
wuhan	panicbuyinguk	pandemic
quarantine	wuflu	quarantineandchill
coronapocalypse	ncov19	wuhancoronavirus

Table 3.1: A partial list of trending keywords used to filter the 1%-Twitter-stream.

Region	Pattern
USA	`SA | SN | usa | united states`
UK	`UK | united kingdom | england | wales | scotlad | northern ireland | britain | london`
Singapore	`SG | sgp | singapore | singapura | 新加坡 | சிங்கப்பூர்`
New Zealand	`new zealand | NZ | aotearoa | wellington | auckland | christchurch | tauranga`

Table 3.2: Full list of expressions used to sort for country from the noisy location tags. `SA` stands for State Abbreviation (i.e. the two letter shortening of each state in the USA) and `SN` stands for State Name.

3.2 Account Preprocessing

To discard extraneous information, standardise formats, and to be able to quickly process accounts, a pre-processing step was performed. First, user accounts with no recorded tweets in the database or with an unspecified location tag were removed from the data set. Then the remaining accounts were sorted to one of the four selected countries based on their location tag. However, as this tag is a free-text string, it was checked whether it contains one of the substrings shown in Table 3.2 to determine which country the account came from. The national-flag-emojis of each country were also checked.

After location based filtering of accounts, the account survival status was determined by querying the Twitter API for each account individually using their unique account ID. Due to limitations of the Twitter API, this process took four days to complete.

The age of accounts was also calculated. This value shows the number of days between the creation date of the account and a selected reference date, which fell to October 24, 2020 in this thesis. This definition of age is standard in the literature [14, 57].

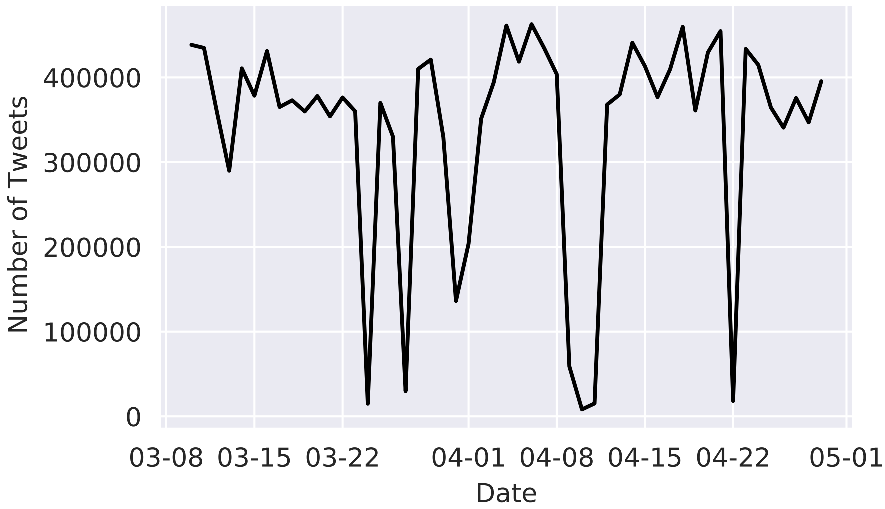


Figure 3.1: Number of tweets in the data set on each day after preprocessing.

Additionally, account reputation [40] was also calculated, which is usually defined as:

$$R = \frac{\#followers}{\#followers + \#friends}$$

In addition to age and reputation, other measures were also calculated, that were previously shown to be effective signals for predicting different account survival types, such as the character-count of the username or the word-count of the biography [57].

3.3 Tweet Preprocessing

After account preprocessing all tweets that were created by one of the remaining accounts in the data set were extracted. The number of tweets on each day is shown in Figure 3.1. Due to outages on Twitter and its API, as well as technical difficulties with the scraping equipment, some of the days have substantially less data than the average.

This was followed by standard text preprocessing steps to normalize tweets for further processing and to remove unnecessary linguistic variations. Tweets were preprocessed using the NLTK package [10] for Python and the ”text” field of tweets. The preprocessing steps are:

- Tokenisation: We used the TweetTokenizer class from NLTK that leaves hashtags, mentions and emoticons intact.
- Case folding: We used Python’s built-in case folding method for its string class.
- Stopword removal: We removed frequently occurring stopwords using a publicly available list[5]. We also removed symbols, punctuation, and the short token ”rt”, which occurs in every retweet.
- Stemming: We used the SnowballStemmer class for English from NLTK.

[5]http://members.unine.ch/jacques.savoy/clef/index.html

3.4 COVID-19 Mentions

For the mention analysis of COVID-19, a range of keywords and account names were selected based on the scraped data set and previous topic models of COVID-19 related tweets [62]. The keywords were collected from a wide range of topics to gauge the overall engagement with COVID-19. For each keyword in the list below, the occurrences of hashtags using the same keyword were also counted. The listed accounts correspond to governmental or regional health agencies.

- Keywords: covid19, coronavirus, corona, ncov, pandemic, epidemic, disease, infection, flu, virus, symptom, outbreak, spread, mask, sanitize, positive, negative, testing, patient, wuhan, case, death, panic, lockdown, quarantine, closure, distancing, health,
- Accounts: @CDCGov, @CDCemergency, @HHSGov, @WHO, @NIH, @NHSuk, @NHSengland, @PublicHealthW, @NHSNSS, @NHS_Lothian, @publichealthni, @sporeMOH, @minhealthnz

To extract how often people tweeted with these keywords, the tweets were grouped by creation day (10/03-29/04), country, and account survival type and then the occurrence counts of the selected keywords and accounts were summed together. Finally, the raw occurrence count was normalised with the total token count of the respective group to obtain *COVID-19 mention frequency*.

3.5 Case Correlations

Correlations using the Pearson correlation-coefficient between daily mention frequencies of COVID-19 and daily case counts were also calculated. Given two sets of data points $x_{1:n}, y_{1:n}$ with means and standard deviations μ, s this is defined as:

$$r_{xy} = \frac{1}{n-1} \sum_{t=1}^{n} \left(\frac{x_t - \mu_x}{s_x} \right) \left(\frac{y_t - \mu_y}{s_y} \right)$$

To assess how case counts may affect interest in COVID-19 on Twitter, the time-lagged Pearson coefficients (TLPC) [50] between mention frequency and case counts were plotted.

TLPC was calculated in each country by keeping case-counts fixed and shifting mention frequencies of COVID-19 in time by an offset of $\tau \in \{-14, -13, \ldots, 14\}$. The TLPC at τ (TLPC[τ]) is then simply given by the Pearson-coefficient between case-counts and the shifted mention frequencies.

Therefore, a TLPC at a *positive offset* τ means that a change (increase if TLPC[τ] > 0 or decrease if TLPC[τ] < 0) in mention frequencies *preceded* a rise in case-counts, while a TLPC at a *negative offset* means that the change in mention frequencies *followed* the rise in case-counts. An example of TLPC is shown in Figure 3.2.

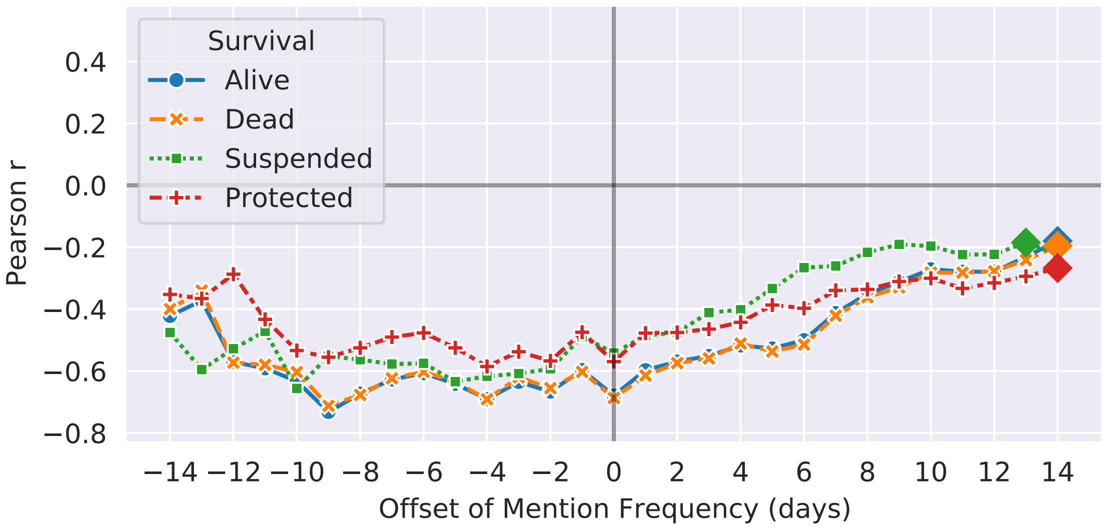


Figure 3.2: Example of TLPC between COVID-19 mention frequencies and case counts in the USA. The highest correlation coefficient is at $\tau = 14$, so mentions of COVID-19 may have preceded actual case-counts by at least two weeks.

Chapter 4

Results

This section presents the results of the thesis divided into two sections. First, the overall results on account survival is presented using primarily account-related metadata and tweeting patterns. These observations are then connected to the results of COVID-19 mention frequency analysis to establish differences in engagement across survival types. Each section is further divided into subsections depending on the variables or topics under investigation.

4.1 Account Survival Analysis

4.1.1 Proportion of Account Survival Types

As Figure 4.1 shows, the distribution of AS types is similar across countries. Of all accounts, alive ones make up $78.85 \pm 3.13\%$, dead $17.23 \pm 1.98\%$, suspended $0.56 \pm 0.17\%$, and protected $3.36 \pm 2.04\%$. There is a only a very small proportion of suspended accounts in the dataset, though their numbers are comparable in proportion to datasets of other studies [14, 53, 20]. Around 15% of all accounts were deleted from Twitter in each country, and there was a larger proportion of protected accounts in Singapore than in other countries, which suggests that people may be more concerned with their accounts' privacy there.

4.1.2 Age

We can observe a very similar age distribution across countries as shown in Figure 4.2. Alive and protected accounts are most similar in age, which highlights that accounts from these types likely come from similar user demographics. Dead accounts on the other hand are significantly younger, with some notable outliers. This suggests that dead accounts, which comprise around 15% of all accounts, might be created with the intention to be thrown-away. The small percentage of outliers could suggest that there are qualitatively different causes behind the deletion of younger and older accounts. Finally, suspended accounts cover a large range of ages, but their distribution aligns closely to observations from other studies [14, 38].

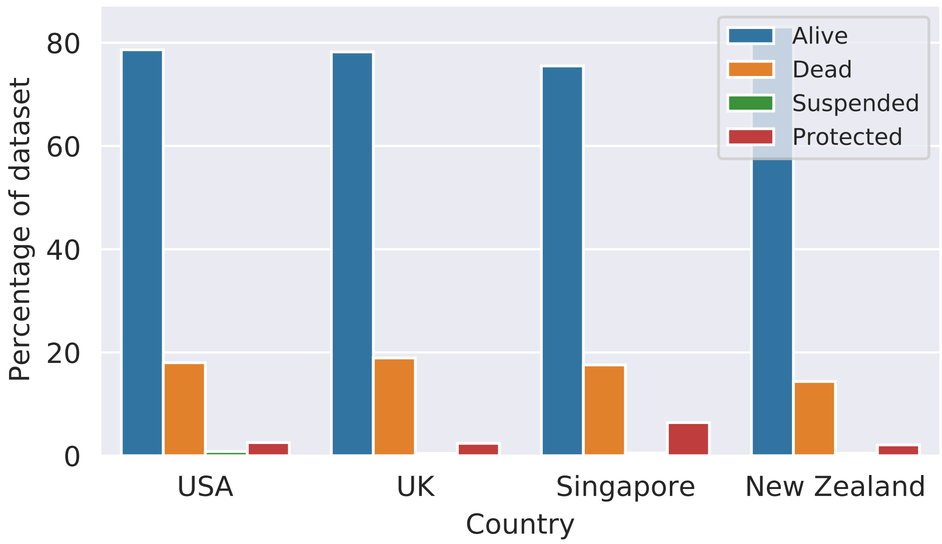


Figure 4.1: Frequency of accounts types in each country.

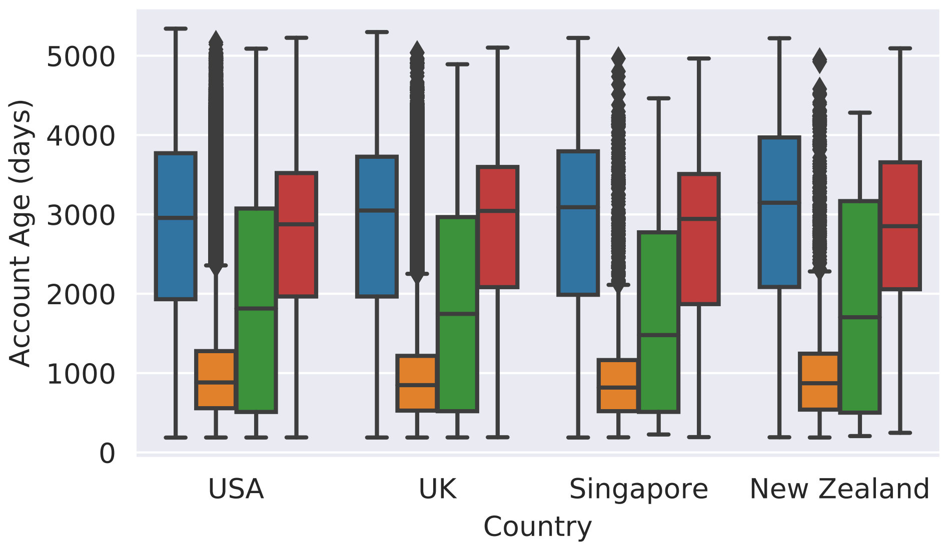


Figure 4.2: Distribution of the age of accounts types in each country.

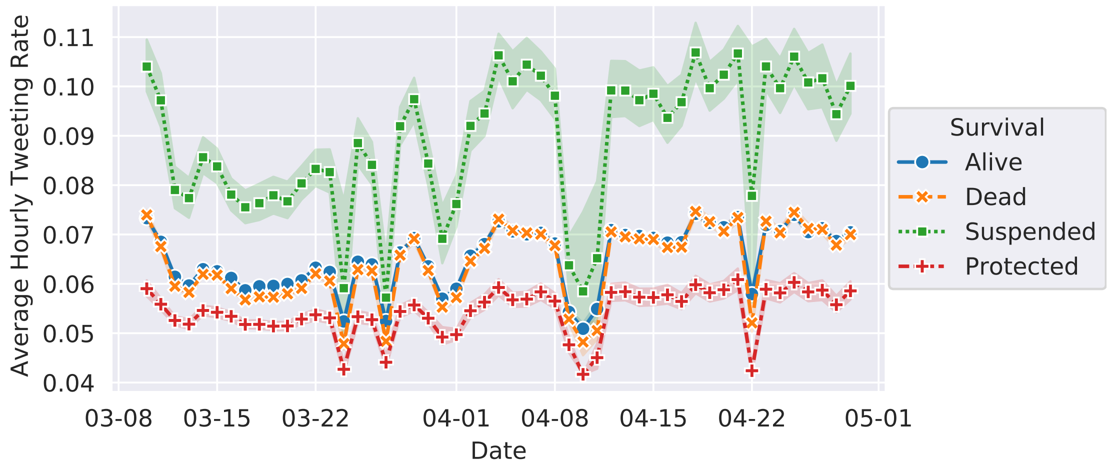


Figure 4.3: Average number of tweets in an hour.

4.1.3 Tweeting Patterns

A wide range of measures were investigated to establish the tweeting behaviours of different AS types. This included for example the number of emojis used, the length of usernames, the rate of URLs in tweets, or the mean number of received favourites. However, most of these measures showed no or insignificant correlation with survival types. Therefore, in this subsection only two relevant and interesting feature are presented. Data was aggregated across countries as the patterns in these features did not significantly change among regions.

The average number of tweets per hour in the dataset broken down for each account survival type is shown in Figure 4.3 with a 95% confidence interval. Suspended accounts tweeted the most often across the examined time period, while protected accounts were the least active. Alive and dead accounts followed almost exactly the same tweeting patterns. This may suggest the existence of similarities between the two AS types, which is in contrast with other findings of this section showing that alive and protected accounts are similar.

In Figure 4.4, the average ratio of retweets to all tweets for an account is showed. Here we can observe that protected accounts issued the most retweets by a significant margin, which could suggest that these accounts post less original content than others. The similarity between alive and dead accounts holds for the retweet-ratio as well. The overall graph shows a slight declining trend, while suspended accounts exhibit a stagnating ratio of retweets to tweets.

4.1.4 Friends and Followers

Figure 4.5 plots the number of friends against the number of followers on a log-scale using a sample size of 2,000. It aggregates accounts across all countries as we observed the same patterns in each country. The cut-off occurring at ~ 8.5 (5,000 friends) is due to a hard limit imposed by Twitter[1]. The solid black line is the identity line of equal

[1] https://help.twitter.com/en/using-twitter/twitter-follow-limit

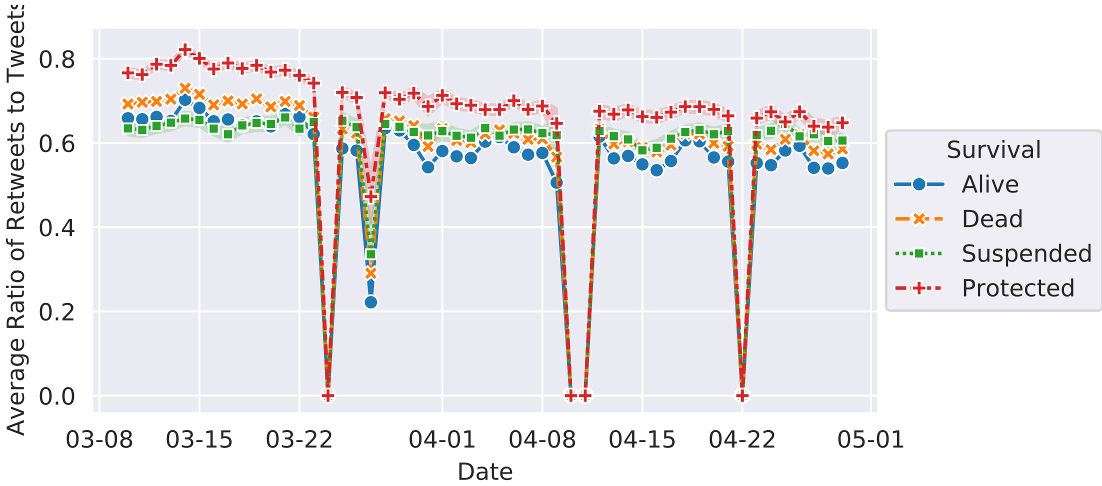


Figure 4.4: The average ratio of retweets to all tweets.

number of friends and followers. The dashed line shows a sixth-degree-polynomial trend-line with a confidence interval of 95%.

We can see the previously observed similarity between alive and protected accounts, as both types follow similar friend-follower distributions. However, more alive accounts have a larger number of followers than friends, because most of the outlying alive accounts are tied to corporate or celebrity users and very few such users protect their accounts. This fact was also already observed in previous studies [54]. Suspended accounts are skewed towards more friends than followers, which is characteristic of spamming accounts [54], but a similar trend is also exhibited by dead accounts. However, dead accounts were more likely to have no followers but a lot of friends.

4.1.5 Reputation and Age

Figure 4.6 shows a probability density estimation calculated using kernel density estimation (KDE) [46] of an account occurring given its reputation and age. Most alive and protected accounts are grouped near a reputation of $R = 0.5$, which means they have a similar number of friends and followers. For these accounts, the older an account is, the more likely it is to have lower reputation, with only alive accounts showing a significant probability to have $R > 0.6$. As remarked in Section 4.1.2 dead accounts are young, but show near uniform distribution for reputations between 0.0 and 0.5.

Dead accounts also shows clear outliers grouping around $R = 0.5$, making these accounts more similar to other AS types. Finally, suspended accounts exhibit an "L-shaped" distribution. As could be expected for spammers or bots, most suspended accounts are grouped near $R \approx 0.0$ and are very short lived. This observation aligns with previous research findings [54]. However there is a significant grouping of older accounts near $R \approx 0.5$, which could correspond to compromised accounts or social bots [14], that masquerade as humans and so managed to avoid detection by automated measures.

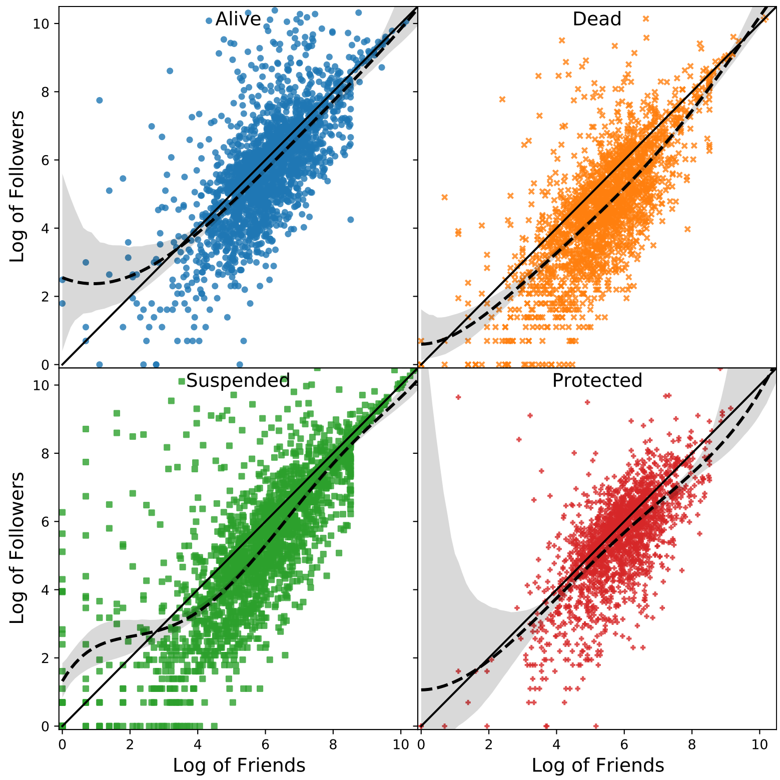


Figure 4.5: Distribution of the number of friends versus followers.

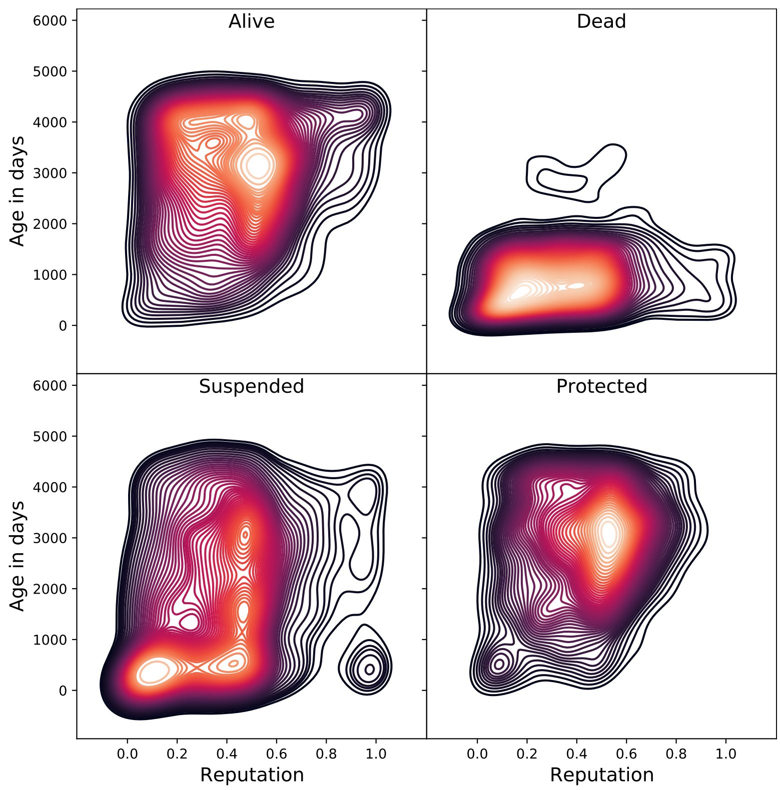


Figure 4.6: Probability density estimation of an account occurring given its reputation and age.

4.2 COVID-19 and Account Survival

4.2.1 COVID-19 Mentions

In Figure 4.7 the evolution of mention frequencies of COVID-19-related keywords for each country and account survival is plotted. In all countries we can observe a clear separation of trend lines between AS types and an overall downward trend in the mention frequencies of keywords, which suggests that the interest in COVID-19 is largely unaffected by case-counts.

In general, protected accounts seem most engaged with COVID-19 in all countries, with the exception of New Zealand which showed significant variability over time across survival types.

In the UK and the US, all four account types show broadly similar patterns of mentions: mostly decreasing mentions of COVID-19 as time progresses.

However, similar patterns do not clearly hold in Singapore and New Zealand. Suspended accounts from Singapore mentioned COVID-19 far less than other account survival types. Dead accounts are also somewhat less likely to mention COVID-19, especially after a spike in cases that coincided with systematic testing in migrant worker dormitories.

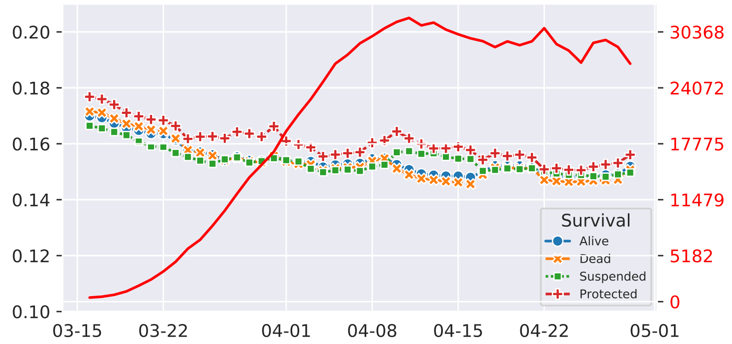
New Zealand shows the least regular patterns. Here, the mentions of COVID-19 by suspended accounts spike before case rates rise, and remain high until cases begin to ease. New Zealand is also the only country that shows a clear dip in Covid-19 mentions for protected accounts.

4.2.2 Case-Count Correlations

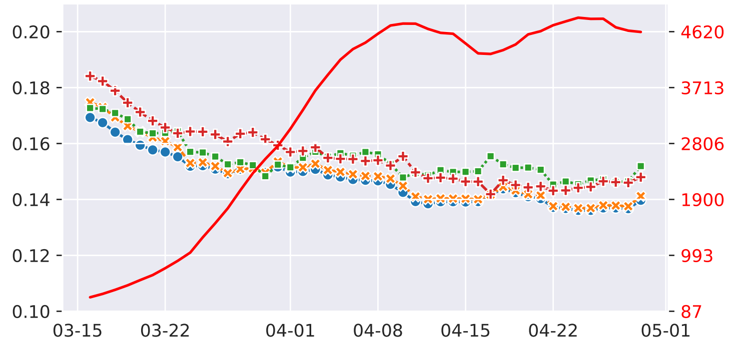
The investigation of the time-lagged Pearson-correlation (TLPC) as described in Section 3.5 between mention frequencies and case-counts is shown in Figure 4.8.

For the USA, UK, and Singapore, almost only negative TLPCs were observed regardless of the offset used and the account survival type. This suggests that interest in COVID-19 was decreasing irrespective of new cases. For the UK and Singapore, correlations tend to be weaker (correlation coefficient closer to 0) for suspended accounts, which means that mentions of COVID-19 are least strongly linked with case counts for this account type.

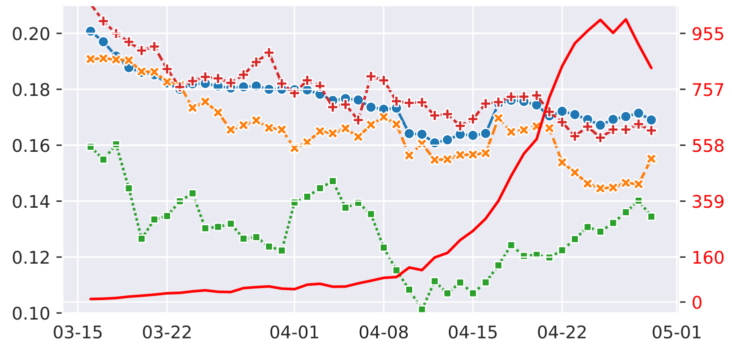
The pattern for New Zealand is different. This was the only country where strong positive correlations were found, though case-counts also began rising here earlier than in other countries. It was also the only country where COVID-19 mentions tracked actual case counts, especially for dead and alive accounts.



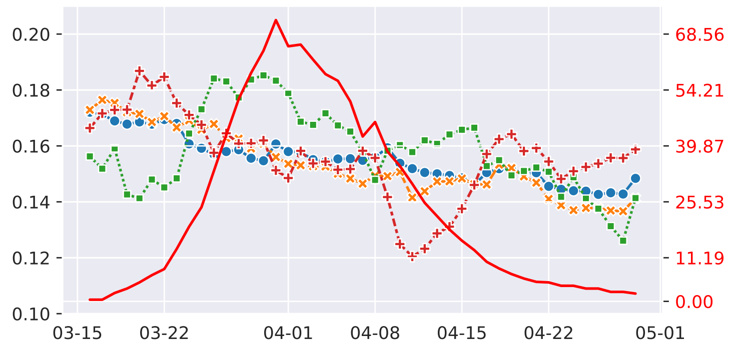
(a) USA



(b) United Kingdom

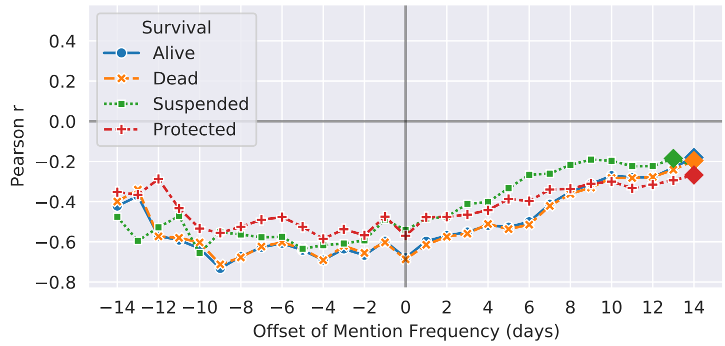


(c) Singapore

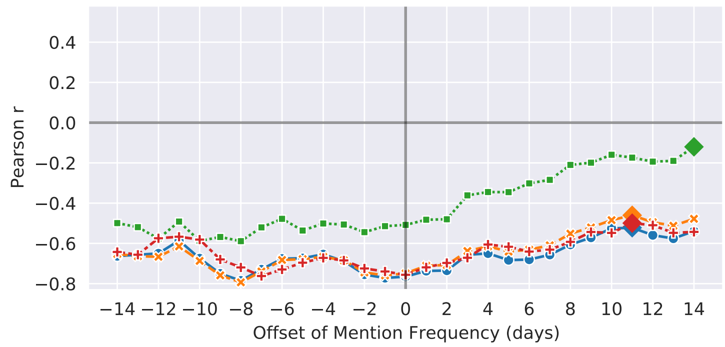


(d) New Zealand

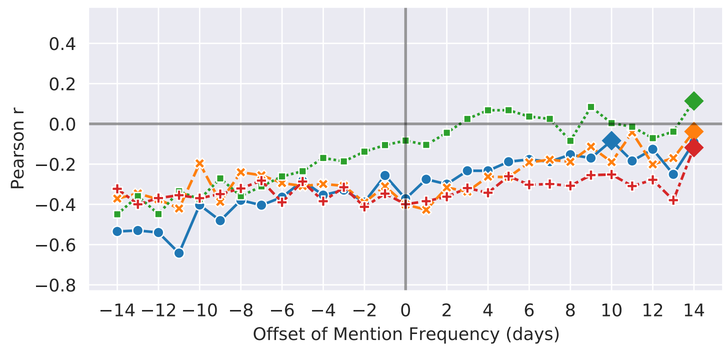
Figure 4.7: 7-day moving average of COVID-19-related mention frequencies for each country. The red line shows the 7-day moving average of new COVID-19 cases in that country.



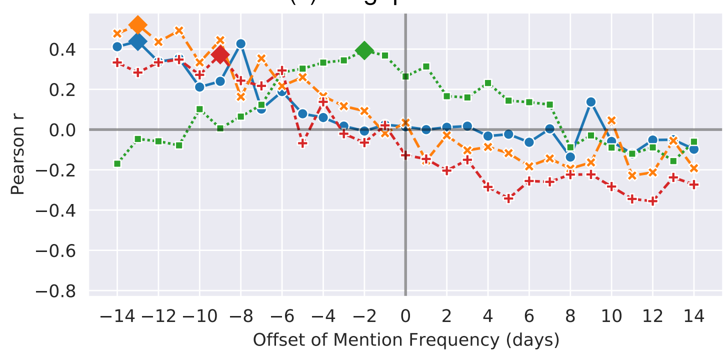
(a) USA



(b) United Kingdom



(c) Singapore



(d) New Zealand

Figure 4.8: Time-lagged Pearson-correlation [50] between mention frequencies and COVID-19 case-counts. Large diamonds mark the position of largest correlation.

Chapter 5

Discussion

This chapter gives answers to the research questions outlined in Section 1.3 and presents a discussion of the results presented in the previous chapter. It also provides a summary of the limitations of the thesis and describes interesting perspectives for future work.

Overall, we can say that the four account survival types behaved according to expectation from the literature. In addition, the comparison of account survival types on a country level provided an opportunity to map the evolution of COVID-19 to changes in social media behaviour.

5.1 RQ1: Characteristics of Account Survival

One of the main results of this thesis was given in Section 4.1, which showed that accounts differ significantly depending on their account survival type.

Age, reputation, and tweeting and retweeting activity were amongst the most important distinguishing features among account survival types, and the behaviour of each account type was as we would expect from the literature. The effects of each of these features is discussed in the following subsections.

The distribution of account survival types was similar across all countries, which suggests that the differences in COVID-19 mentions between countries is not due to differences in the number of relevant accounts. Instead, there could exist latent factors that produced these differences among countries.

5.1.1 Age

In Section 4.1.2, we saw, that the age distributions of alive and protected accounts were very similar. This highlights that their underlying user base could be very similar.

The young age of dead accounts hint that these were made as throw-away accounts, however the outliers here point to a qualitatively different subgroup of dead accounts, which could be investigated in-depth.

Finally, the large range of ages of suspended accounts aligns with previous findings [14]. A possible reason for this wide age-interval might be the co-existence of social bots and traditional bots in the Twitter-sphere, as the former could evade detection for much longer than the latter.

5.1.2 Tweeting Frequency and Retweet Ratio

According to the results of Section 4.1.3, suspended accounts issued on average the most number of tweets, but showed the lowest retweet-ratio. The former observation is a characteristic often mentioned in connection to spamming accounts. On the other hand, the latter observation is in contrast with previous findings on automated accounts [39]. This could suggest that suspended accounts in this data set were more likely to be spammers than bots.

Protected accounts were the least active tweeters but had the highest retweet-ratio, which means they issued the least amount of original content. Further investigation of protected accounts is an important next step for the understanding of this phenomenon.

Finally, alive and dead accounts displayed very similar tweeting trends. This sets alive and protected accounts apart which is in contrast with their previously observed similarities.

5.1.3 Reputation

The low reputation of dead accounts provides evidence that these accounts are made to be thrown away, though the outliers, similarly to the age distribution, form a separate group necessary of investigation.

The "L-shaped" distribution of suspended accounts also shows the existence of multiple user groups in this AS type. Most young accounts with near-zero reputation are likely to be bots, however accounts grouped around $R \approx 0.5$ may have instead been suspended due to security reasons. They may also be social spambots [14] who masquerade as humans.

Finally, dead accounts which have no followers, but a lot of friends, might be fake-followers [14] that were created to artificially inflate the number of followers of an account.

5.2 RQ2: Mentions of COVID-19

While there was a clear separation of account survival types when looking at the mention frequencies of COVID-19, the actual mention patterns differed strongly by country. Accounts in the USA and UK followed similar trends with protected accounts being most interested in COVID-19 for the majority of the observed time period. A steady decrease in the mentions of COVID-19 could also clearly be observed.

On the other hand, New Zealand showed a lot of variation over the studied time-period. Despite the early ending of the pandemic, mention frequencies did not subside. Overall, suspended accounts had the lowest mention frequencies of COVID-19. If there had been an increase in COVID-19 related spambots in the early days of the pandemic, we would have expected to see the opposite effect. Suspended accounts in Singapore even appeared to be actively avoiding the topic of the pandemic, with very low mention frequencies compared to the other account types.

For all countries, protected accounts tended to mention COVID-19 slightly more frequently than others. However, all that is known about these protected accounts is that they were public during the first wave of COVID-19. It is not known at which point in time they restricted access, so we cannot conclude what may have triggered the decision to protect their accounts, and whether it was related to the pandemic.

5.3 RQ3: Effect of New Cases of COVID-19

In most of the countries studied, the number of mentions of COVID-19 was negatively correlated with the case count. The pandemic was discussed mostly towards the start, in March 2020. The exception was New Zealand, where online discussion mirrored case counts.

This negative correlation and the results of Section 4.2.2 suggest, that interest in the pandemic on Twitter was not directly related to the incidence of new cases of COVID-19.

5.4 Limitations

No gold standard was available for location filtering, or geo-tagging, and user-provided information was used instead. While this introduces a substantial amount of additional noise into the data set, it respects users' rights to determine the location they wish to be associated with.

Additionally, the thesis did not account for changes in the definitions of case counts across countries, nor did it account for uncertainty in case counts.

In the US and the UK, there was additional substantial variation at the state level (US) and at the level of devolved governments (UK) which was not taken into account.

A further limitation is presented in the form of missing data, which was due to technical problems with scraping Twitter. However, the normalisation of the statistics and mention frequencies guaranteed that the effect of differing absolute sizes was minimised.

For the US and Singapore, the data set was further limited by the restriction to English language tweets. While English is the lingua franca of Singapore, and the main language of the US, we would also expect substantial activity in Spanish (US) and Mandarin / Hokkien, Tamil, and Malay (Singapore).

Finally, given that the data set was sampled from the Twitter streaming API, which only covers 1% of all traffic, this may further bias the sample of the analysis.

Chapter 6

Conclusions

This chapter gives a summary of the background, and methods presented in the thesis, and provides an overview of the main results of this work. The chapter and the thesis is concluded with a discussion of further research directions, that could build on the contents of this project.

6.1 Summary

This thesis presented a detailed, quantitative comparison of user account survival on Twitter during the first wave of COVID-19. The comparison was performed based on a four-class categorisation, which assigned accounts into exactly one of the following four survival types: alive, dead, suspended, and protected. The protected survival type is a novel addition of the thesis.

As the management of COVID-19 showed great variance across regions of the world, the thesis further investigated account survival using a cross-country comparison among four nations: the United States of America, United Kingdom, Singapore, and New Zealand.

Account survival is a relatively understudied area of social media analysis. However, better understanding of the factors affecting what drives users to change their accounts could, in general, provide important insights for several related fields, such as online privacy, social media (dis)continuance, and spamming. Additionally, in the context of COVID-19, online user behaviour was shown to have changed significantly. This work therefore also studies the extent to which account survival has been affected by the pandemic and the incidence of new cases.

6.1.1 Results

The results showed significant, quantitative differences among survival types, however case-counts had negligible or no effects on the interest in COVID-19.

The overall patterns of account survival stayed similar across the four countries. Suspended accounts were seen to be most active in terms of the number of tweets issued, while protected accounts were the least active tweeters. Dead and alive accounts showed remarkably close trends, which hints that these accounts may share a common user demographic. On the other hand, protected accounts have produced the most number of retweets on average. This suggests, that these accounts were issuing less original tweets than other survival types, which may be an early sign of attention avoidance or a method of privacy self-preservation.

In the US and the UK, the mention patterns of COVID-19 followed similar trends, regardless of account type. We can observe an overall decreasing trend in the mention frequencies, which shows that case-counts did not affect interest in the pandemic. Furthermore, New Zealand showed substantial variation over time, however the early end of the pandemic better aligned case-counts with the evolution of mention frequencies. Finally, suspended accounts from Singapore were less likely to mention COVID-19 than all other account types, which could however be a general trend in Singapore and not particular to COVID-19.

We can conclude that account survival types should be taken into consideration when we investigate how the COVID-19 pandemic is discussed on Twitter.

6.2 Further Work

The present work provides a number of interesting trends through quantitative observations, that could form the basis for several further studies. In the following, I give five major directions that could be pursued in the future.

First, the trends of this thesis should be contrasted with information from a baseline data set, without filtering of any kind. This would allow us to differentiate between COVID-19 specific trends and overall account survival patterns. A substantial enhancement to the presented methods would be to study the temporal evolution of account survival, instead of a single, one-time check. This approach, though more computationally expensive, could uncover correlations between real-life events and account survival on Twitter.

Second, account survival types show distinct patterns of behaviour, which warrants a closer inspection of their behaviour. The methods for this could involve a range of established techniques that study the behaviour of social media users similarly to the linguistic and network analyses discussed in Section 2.2. Some of the patterns, that could be worth understanding through qualitative analyses are:

- Why do suspended accounts exhibit and ”L-shaped” distribution of account age against reputation?
- What accounts constitute the outliers of the dead survival type and what are their main characteristics?
- Why do we observe close similarities in the tweeting patterns between alive and dead accounts?

Third, the apparent lack of effect of the incidence of COVID-19 on mention frequencies is surprising, and could be the topic of further study. This line of analysis is especially interesting in the case of suspended accounts. As these accounts are often associated with bots and spammers, and the pandemic founded a basis of an ”infodemic”, it is worth examining the most frequently discussed topics of suspended accounts, as well as the temporal and spatial distribution of their tweets.

Fourth, the thesis showed that the so far understudied protected survival type is qualitatively different from other accounts. It is therefore beneficial to investigate these accounts more in depth. For instance, looking at why protected accounts have the highest retweet-to-tweet ratio but the lowest tweeting frequency could uncover some of the driving causes of account protection. The similarity in age and reputation distribution with alive accounts could provide further clues into the causes behind protection. Protected accounts also showed greater interest in COVID-19. The study of this phenomenon could shed light on people’s privacy concerns during the pandemic.

Finally, the spatial analysis of account survival should be extended, and other socio-economic factors, such as race, gender, and financial standing could be differentiated among survival types. Though, inferring socio-economic factors from accounts is a difficult task which also raises various ethical issues, however involving such elements could shed light on the social media usage of various subgroups of people and provide helpful insights into tackling digital exclusion or bias.

Bibliography

[1] A. H. Alamoodi, B. B. Zaidan, A. A. Zaidan, O. S. Albahri, K. I. Mohammed, R. Q. Malik, E. M. Almahdi, M. A. Chyad, Z. Tareq, A. S. Albahri, Hamsa Hameed, and Musaab Alaa. Sentiment analysis and its applications in fighting COVID-19 and infectious diseases: A systematic review. *Expert Systems with Applications*, page 114155, October 2020.

[2] Hazim Almuhimedi, Shomir Wilson, Bin Liu, Norman Sadeh, and Alessandro Acquisti. Tweets are forever: A large-scale quantitative analysis of deleted tweets. *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*, February 2013.

[3] Christina J Atchison, Leigh Bowman, Charlotte Vrinten, Rozlyn Redd, Philippa Pristerà, Jeffrey W Eaton, and Helen Ward. Perceptions and behavioural responses of the general public during the covid-19 pandemic: A cross-sectional survey of uk adults. *medRxiv*, 2020.

[4] Isabelle Augenstein, Tim Rocktäschel, Andreas Vlachos, and Kalina Bontcheva. Stance detection with bidirectional conditional encoding. *arXiv preprint arXiv:1606.05464*, 2016.

[5] Ramakrishna Ayyagari, Varun Grover, and Russell Purvis. Technostress: Technological Antecedents and Implications. *MIS Quarterly*, 35(4):831–858, 2011. Publisher: Management Information Systems Research Center, University of Minnesota.

[6] Randy Baden, Adam Bender, Neil Spring, Bobby Bhattacharjee, and Daniel Starin. Persona: an online social network with user-defined privacy. In *Proceedings of the ACM SIGCOMM 2009 conference on Data communication*, pages 135–146, 2009.

[7] Joseph A Balogun. Commentary: Lessons from the usa delayed response to the covid-19 pandemic. *African journal of reproductive health*, 24(1):14–21, 2020.

[8] Susanne Barth, Menno D. T. de Jong, Marianne Junger, Pieter H. Hartel, and Janina C. Roppelt. Putting the privacy paradox to the test: Online privacy and security behaviors among users with technical knowledge, privacy awareness, and financial resources. *Telematics and Informatics*, 41:55–69, 2019.

[9] Marco T Bastos. This account doesn't exist: Tweet decay and the politics of deletion in the brexit debate. *Available at SSRN 3485789*, 2019.

[10] Steven Bird, Ewan Klein, and Edward Loper. *Natural language processing with Python: analyzing text with the natural language toolkit.* " O'Reilly Media, Inc.", 2009.

[11] Farhan Asif Chowdhury, Lawrence Allen, Mohammad Yousuf, and Abdullah Mueen. On Twitter Purge: A Retrospective Analysis of Suspended Users. In *Companion Proceedings of the Web Conference 2020*, WWW '20, pages 371–378, New York, NY, USA, April 2020. Association for Computing Machinery.

[12] Zi Chu, Indra Widjaja, and Haining Wang. Detecting Social Spam Campaigns on Twitter. In Feng Bao, Pierangela Samarati, and Jianying Zhou, editors, *Applied Cryptography and Network Security*, Lecture Notes in Computer Science, pages 455–472, Berlin, Heidelberg, 2012. Springer.

[13] Matteo Cinelli, Walter Quattrociocchi, Alessandro Galeazzi, Carlo Michele Valensise, Emanuele Brugnoli, Ana Lucia Schmidt, Paola Zola, Fabiana Zollo, and Antonio Scala. The COVID-19 social media infodemic. *Scientific Reports*, 10(1), 10 2020.

[14] S. Cresci, R. D. Pietro, M. Petrocchi, A. Spognardi, and M. Tesconi. Social Fingerprinting: Detection of Spambot Groups Through DNA-Inspired Behavioral Modeling. *IEEE Transactions on Dependable and Secure Computing*, 15(4):561–576, July 2018.

[15] Jose Yunam Cuan-Baltazar, Maria José Muñoz-Perez, Carolina Robledo-Vega, Maria Fernanda Pérez-Zepeda, and Elena Soto-Vega. Misinformation of COVID-19 on the Internet: Infodemiology Study. *JMIR Public Health and Surveillance*, 6(2):e18444, 2020. Company: JMIR Public Health and Surveillance Distributor: JMIR Public Health and Surveillance Institution: JMIR Public Health and Surveillance Label: JMIR Public Health and Surveillance Publisher: JMIR Publications Inc., Toronto, Canada.

[16] Leucio Antonio Cutillo, Refik Molva, and Thorsten Strufe. Safebook: A privacy-preserving online social network leveraging on real-life trust. *IEEE Communications Magazine*, 47(12):94–101, 2009.

[17] Susanna Every-Palmer, Matthew Jenkins, Philip Gendall, Janet Hoek, Ben Beaglehole, Caroline Bell, Jonathan Williman, Charlene Rapsey, and James Stanley. Psychological distress, anxiety, family violence, suicidality, and wellbeing in new zealand during the covid-19 lockdown: A cross-sectional study. *PLoS one*, 15(11):e0241658, 2020.

[18] Shaoxiong Fu, Hongxiu Li, Yong Liu, Henri Pirkkalainen, and Markus Salo. Social media overload, exhaustion, and use discontinuance: Examining the effects of information overload, system feature overload, and social overload. *Information Processing & Management*, 57(6):102307, 2020.

[19] Junling Gao, Pinpin Zheng, Yingnan Jia, Hao Chen, Yimeng Mao, Suhong Chen, Yi Wang, Hua Fu, and Junming Dai. Mental health problems and social media exposure during covid-19 outbreak. *PLOS ONE*, 15(4):1–10, 04 2020.

[20] Chris Grier, Kurt Thomas, Vern Paxson, and Michael Zhang. @spam: The underground on 140 characters or less. In *Proceedings of the 17th ACM Conference on Computer and Communications Security*, CCS '10, page 27–37, New York, NY, USA, 2010. Association for Computing Machinery.

[21] Balint Gyevnar. Using usernames to predict the survival of accounts on Twitter, 2020.

[22] Hana Habib, Sarah Pearman, Jiamin Wang, Yixin Zou, Alessandro Acquisti, Lorrie Faith Cranor, Norman Sadeh, and Florian Schaub. "It's a scavenger hunt": Usability of Websites' Opt-Out and Data Deletion Choices. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, pages 1–12, New York, NY, USA, April 2020. Association for Computing Machinery.

[23] Hana Habib, Yixin Zou, Aditi Jannu, Neha Sridhar, Chelse Swoopes, Alessandro Acquisti, Lorrie Faith Cranor, Norman Sadeh, and Florian Schaub. An Empirical Analysis of Data Deletion and Opt-Out Choices on 150 Websites. 2019.

[24] Joanne Hinds, Emma J. Williams, and Adam N. Joinson. "It wouldn't happen to me": Privacy concerns and perspectives following the Cambridge Analytica scandal. *International Journal of Human-Computer Studies*, 143:102498, 2020.

[25] Hillbun Ho and Kenichi Ito. Consumption-oriented engagement in social network sites: Undesirable influence on personal well-being. *European Journal of Marketing*, 53(7):1355–1377, January 2019. Publisher: Emerald Publishing Limited.

[26] David J. Hunter. Covid-19 and the stiff upper lip — the pandemic response in the united kingdom. *New England Journal of Medicine*, 382(16):e31, 2020.

[27] Alexa Internet Inc. twitter.com Competitive Analysis, Marketing Mix and Traffic - alexa. `https://www.alexa.com/siteinfo/twitter.com`. Accessed: 2021-03-31.

[28] A. K. M. Najmul Islam, Samuli Laato, Shamim Talukder, and Erkki Sutinen. Misinformation sharing and social media fatigue during COVID-19: An affordance and cognitive load perspective. *Technological Forecasting and Social Change*, 159:120201, October 2020.

[29] Sarah Jefferies, Nigel French, Charlotte Gilkison, Giles Graham, Virginia Hope, Jonathan Marshall, Caroline McElnay, Andrea McNeill, Petra Muellner, Shevaun Paine, et al. Covid-19 in new zealand and the impact of the national response: a descriptive epidemiological study. *The Lancet Public Health*, 5(11):e612–e623, 2020.

[30] Garrett A. Johnson, Scott K. Shriver, and Shaoyin Du. Consumer Privacy Choice in Online Advertising: Who Opts Out and at What Cost to Industry? *Marketing Science*, 39(1):33–51, 2020. Publisher: INFORMS.

[31] Leo H Kahane. Politicizing the mask: Political, economic and demographic factors affecting mask wearing behavior in the usa. *Eastern Economic Journal*, pages 1–21.

[32] Kawaljeet Kaur Kapoor, Kuttimani Tamilmani, Nripendra P Rana, Pushp Patil, Yogesh K Dwivedi, and Sridhar Nerur. Advances in social media research: Past, present and future. *Information Systems Frontiers*, 20(3):531–558, 2018.

[33] Matthew D. Kearney, Shawn C. Chiang, and Philip M. Massey. The Twitter origins and evolution of the COVID-19 "plandemic" conspiracy theory. *Harvard Kennedy School Misinformation Review*, 1(3), October 2020.

[34] Dilara Keküllüoglu, Walid Magdy, and Kami Vaniea. Analysing privacy leakage of life events on twitter. In *12th ACM Conference on Web Science*, WebSci '20, page 287–294, New York, NY, USA, 2020. Association for Computing Machinery.

[35] Efthymios Kouloumpis, Theresa Wilson, and Johanna Moore. Twitter sentiment analysis: The good the bad and the omg! In *Fifth International AAAI conference on weblogs and social media*. Citeseer, 2011.

[36] Ramez Kouzy, Joseph Abi Jaoude, Afif Kraitem, Molly B El Alam, Basil Karam, Elio Adib, Jabra Zarka, Cindy Traboulsi, Elie W Akl, and Khalil Baddour. Coronavirus Goes Viral: Quantifying the COVID-19 Misinformation Epidemic on Twitter. *Cureus*, 12(3), 2020.

[37] Clare Llewellyn, Laura Cram, Adrian Favero, and Robin L. Hill. Russian Troll Hunting in a Brexit Twitter Archive. In *Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries*, JCDL '18, pages 361–362, New York, NY, USA, 2018. ACM. event-place: Fort Worth, Texas, USA.

[38] Clare Llewellyn, Laura Cram, Robin L Hill, and Adrian Favero. For whom the bell trolls: Shifting troll behaviour in the twitter brexit debate. *JCMS: Journal of Common Market Studies*, 57(5):1148–1164, 2019.

[39] Michele Mazza, Stefano Cresci, Marco Avvenuti, Walter Quattrociocchi, and Maurizio Tesconi. RTbust: Exploiting Temporal Patterns for Botnet Detection on Twitter. In *Proceedings of the 10th ACM Conference on Web Science*, WebSci '19, pages 183–192, New York, NY, USA, 2019. ACM. event-place: Boston, Massachusetts, USA.

[40] M. McCord and M. Chuah. Spam detection on twitter using traditional classifiers. In Jose M. Alcaraz Calero, Laurence T. Yang, Félix Gómez Mármol, Luis Javier García Villalba, Andy Xiaolin Li, and Yan Wang, editors, *Autonomic and Trusted Computing*, pages 175–186, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.

[41] Brad A. Meisner. Are you ok, boomer? intensification of ageism and intergenerational tensions on social media amid covid-19. *Leisure Sciences*, pages 1–6, 2020.

[42] Hiroyoshi Morita, Hirokazu Kato, and Yoshitsugu Hayashi. International Comparison of Behavior Changes with Social Distancing Policies in Response to COVID-19. 5 2020.

[43] Teagen Nabity-Grover, Christy M. K. Cheung, and Jason Bennett Thatcher. Inside out and outside in: How the COVID-19 pandemic affects self-disclosure on social media. *International Journal of Information Management*, 55:102188, 2020.

[44] Adrian O Dowd. Covid-19: Uk test and trace system still missing 80% target for reaching contacts. *BMJ*, 370, 2020.

[45] Sangchul Park, Gina Jeehyun Choi, and Haksoo Ko. Information Technology–Based Tracing Strategy in Response to COVID-19 in South Korea—Privacy Controversies. *JAMA*, 323(21):2129, June 2020.

[46] Emanuel Parzen. On estimation of a probability density function and mode. *Ann. Math. Statist.*, 33(3):1065–1076, 09 1962.

[47] Leonie Reichert, Samuel Brack, and Björn Scheuermann. Privacy-Preserving Contact Tracing of COVID-19 Patients. Technical Report 375, 2020.

[48] Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo. Earthquake shakes twitter users: real-time event detection by social sensors. In *Proceedings of the 19th international conference on World wide web*, pages 851–860, 2010.

[49] Roshan Santosh, H Andrew Schwartz, Johannes Eichstaedt, Lyle Ungar, and Sharath Chandra Guntuku. Detecting Emerging Symptoms of COVID-19 using Context-based Twitter Embeddings. page 7, 2020.

[50] Chenhua Shen. Analysis of detrended time-lagged cross-correlation between two nonstationary time series. *Physics Letters A*, 379(7):680–687, 2015.

[51] Shefaly Shorey, Emily Ang, Amira Yamina, and Clarence Tam. Perceptions of public on the COVID-19 outbreak in Singapore: a qualitative content analysis. *Journal of Public Health*, 42(4):665–671, 07 2020.

[52] P. Staszkiewicz, I. Chomiak-Orsa, and I. Staszkiewicz. Dynamics of the covid-19 contagion and mortality: Country factors, social media, and market response evidence from a global panel analysis. *IEEE Access*, 8:106009–106022, 2020.

[53] Kurt Thomas, Chris Grier, Dawn Song, and Vern Paxson. Suspended accounts in retrospect: an analysis of twitter spam. In *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*, IMC '11, pages 243–258, New York, NY, USA, November 2011. Association for Computing Machinery.

[54] Kurt Thomas, Chris Grier, Dawn Song, and Vern Paxson. Suspended accounts in retrospect: An analysis of twitter spam. In *Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference*, IMC '11, page 243–258, New York, NY, USA, 2011. Association for Computing Machinery.

[55] Sho Tsugawa, Yusuke Kikuchi, Fumio Kishino, Kosuke Nakajima, Yuichi Itoh, and Hiroyuki Ohsaki. Recognizing depression from twitter activity. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, pages 3187–3196, 2015.

[56] Joshua Uyheng and Kathleen M. Carley. Bots and online hate during the COVID-19 pandemic: case studies in the United States and the Philippines. *Journal of Computational Social Science*, October 2020.

[57] Svitlana Volkova and Eric Bell. Identifying Effective Signals to Predict Deleted and Suspended Accounts on Twitter across Languages. 2017.

[58] Marcin Waniek, Tomasz P Michalak, Michael J Wooldridge, and Talal Rahwan. Hiding individuals and communities in a social network. *Nature Human Behaviour*, 2(2):139–147, 2018.

[59] Wei Wei, Kenneth Joseph, Huan Liu, and Kathleen M. Carley. Exploring characteristics of suspended users and network stability on Twitter. *Social Network Analysis and Mining*, 6(1):51, July 2016.

[60] John EL Wong, Yee Sin Leo, and Chorh Chuan Tan. Covid-19 in singapore—current experience: critical global issues that require attention and action. *Jama*, 323(13):1243–1244, 2020.

[61] Elena Zheleva and Lise Getoor. To join or not to join: The illusion of privacy in social networks with mixed public and private user profiles. WWW '09, page 531–540, New York, NY, USA, 2009. Association for Computing Machinery.

[62] Han Zheng, Dion Hoe-Lian Goh, Edmund Wei Jian Lee, Chei Sian Lee, and Yin-Leng Theng. Uncovering Topics Related to COVID-19 Pandemic on Twitter. In Emi Ishita, Natalie Lee San Pang, and Lihong Zhou, editors, *Digital Libraries at Times of Massive Societal Transition*, Lecture Notes in Computer Science, pages 307–312, Cham, 2020. Springer International Publishing.

[63] Lu Zhou, Wenbo Wang, and Keke Chen. Tweet properly: Analyzing deleted tweets to understand and identify regrettable ones. In *Proceedings of the 25th International Conference on World Wide Web*, pages 603–612, 2016.

[64] Michael Zimmer and Nicholas John Proferes. A topology of Twitter research: disciplines, methods, and ethics. *Aslib Journal of Information Management*, 66(3):250–261, January 2014. Publisher: Emerald Group Publishing Limited.