

Social Media and the COVID-19: South African and Zimbabwean Netizens' Response to a Pandemic

M B Mutanga^{*1}, O Ureke², T Chani³

^{1,3}Department of Information and Communication Technology, Mangosuthu University of Technology, Durban, South Africa

²Department of Media and Society Studies, Midlands State University, Gweru, Zimbabwe

E-mail: mutangamb@mut.ac.za¹, urekeo@msu.ac.zw², chani@mut.ac.za³

Submitted: 14 March 2021, revised: 14 July 2021, accepted: 21 July 2021

Abstrak. Sejak akhir tahun 2019, dunia menghadapi krisis kesehatan berupa pandemi virus corona (COVID-19). Untuk mengurangi dampak pandemi, pemerintah di seluruh dunia menerapkan aturan seperti membatasi perjalanan lokal dan internasional, serta memerintahkan warga untuk tinggal di dalam rumah. Dengan mempertimbangkan dampak sosial dan ekonomi dari pembatasan ini, menjadi penting untuk meneliti persepsi warga internet (netizen) tentang berbagai upaya yang diambil. Studi ini berpegang pada teori publik digital, yang memanfaatkan media sosial sebagai *platform virtual* di mana para netizen saling berbagi ide dan perdebatan tentang isu-isu yang mempengaruhi mereka. Namun, sebagian besar data ini tidak terstruktur dan sulit untuk diolah. *Natural language processing (NLP)* mengumpulkan dan menganalisis sejumlah besar data tekstual menjadi layak untuk diolah. Cara kerjanya adalah dengan mengekstraksi pengetahuan terstruktur dari *natural language*. Tentu terdapat tantangan yaitu sifat linguistik yang beragam termasuk singkatan, kesalahan ejaan, tanda baca, kata berhenti, dan teks yang tidak baku. Studi ini menerapkan algoritma *The Latent Dirichlet Allocation (LDA)* pada data Twitter untuk mengekstrak topik yang dibahas oleh netizen dari Zimbabwe dan Afrika Selatan. Fokus studi ini adalah mengeksplorasi berbagai topik yang meramaikan komunitas di Twitter dari kedua negara secara komparatif. Studi ini juga meneliti apakah identitas nasional yang mendefinisikan dan membedakan warga negara-negara ini juga ada di Twitter, serta menyelidiki opini publik dengan menganalisis bagaimana warga mendiskusikan isu seputar pandemi COVID-19 di media sosial.

Kata kunci: ekstraksi topik; *natural language processing*; COVID-19; Zimbabwe; Afrika Selatan

Abstract. Since the end of 2019, the world faced a major health crisis in the form of the Coronavirus (COVID-19) pandemic. To mitigate the impact of the pandemic, governments across the globe instituted measures such as restricting local and international travel and, in many cases, ordering citizens to stay indoors. Considering the social and economic impact of these restrictions it becomes crucial to investigate internet citizens' (netizens) perception about the precautionary measures adopted. The study is anchored in the digital public sphere theory, which treats social media applications as virtual platforms where netizens commune to share ideas and debate about issues that affect them. However, the majority of this data is unstructured and difficult to interpret. Natural language processing (NLP), on the other hand, makes the task of gathering and analysing vast amounts of textual data feasible. Extracting

structured knowledge from natural language, however, comes with unique challenges due to diverse linguistic properties including abbreviation, spelling mistakes, punctuations, stop words and non-standard text. In this work, The Latent Dirichlet Allocation (LDA) algorithm was applied to Twitter data to extract topics discussed by netizens from Zimbabwe and South Africa. The focus of this paper is to comparatively explore the variety of topics that occupied Twitter communities from the two countries. The paper examined whether the national identities that define and differentiate citizens of these countries also exist on Twitter as evident in the emerging topics. Furthermore, this work investigated public opinion by analysing how citizens discuss the issues around the COVID-19 pandemic on social media.

Keywords: topic extraction; natural language processing; COVID-19; Zimbabwe; South Africa

1. Introduction

The Corona virus (COVID-19) that ravaged the world since its emergence in China in December 2019 was declared a pandemic by the World Health Organisation (WHO). The illness affected millions and killed hundred thousand people, while many more lived in fear. To reduce the impact of the pandemic, governments across the globe instituted measures such as restricting local and international travel and, in many cases, ordering citizens to stay at their homes under complete lockdown. At the same time, scientists and pharmaceutical companies started the process of developing either vaccines or a cure. The success of such synergized efforts largely depended on citizens' attitudes and perceptions. Globally, citizens were gripped by panic and anxiety exacerbated by the rapid spread of the disease and the unusual mitigatory measures adopted by governments. Considering the social and economic impact of these restrictions, it becomes critical to investigate internet citizens' (netizens) reactions to the pandemic and the measures instituted by the governments. Such an investigation, among other benefits, may serve as the basis to measure the effectiveness of these measures. This study is anchored in the digital public sphere theory, treating social media applications as virtual platforms where netizens commune to share ideas and debate about issues that affect them. It is evident that social media platforms already carry critical public views on the current pandemic. However, most of this data is unstructured and difficult to analyse using traditional statistical software. Traditionally, surveys are most suitable for such studies aimed at extracting and assessing public opinion. However, in the face of a major disease outbreak that has imposed social isolation, the conventional survey becomes impracticable. This shortcoming, therefore, requires the use of other forms of data gathering tools or methods as proposed in this study.

The primary goal of the paper is to comparatively explore the discourses on a variety of topics that occupied 'Zimbabwean' and 'South African' Twitter communities at the beginning of the COVID-19 crisis. The study is delimited to Tweets posted between March and May 2020, a period during which both South Africa and Zimbabwe were under total lockdown. This work examines the tweets, the fears, frustrations, and speculations of the Zimbabwean and South African Twitter communities based on the emerging topics. It was commonplace where official communication lacked or was mistrusted, and citizens subsequently initiated their own communication to alert each other about realities of the pandemic. In Zimbabwe, for example, netizens convened in a Facebook group called 'Coronavirus COVID-19 Zimbabwe' where information, including official statistics, was posted.

This paper further examines whether the national identities that defined and differentiated citizens of these countries in 'normal' times also existed on Twitter as evident in the topics discussed in their commentary. In recent times, South African citizens have complained against, and at times physically attacked foreign nationals, including Zimbabweans residing in the country, accusing them of depriving them of employment opportunities and committing various crimes. This has seen a surge in xenophobic attacks targeting mainly black non-South Africans. The conflict has, at times, spilt onto social media in the forms of Twitter wars (twars). The current study attempts to find out whether these sharp differences on national grounds continued to characterise Zimbabwe-South Africa discourses in the face of a common pandemic.

South Africa recorded its first case of COVID-19 on the 5th of March while Zimbabwe reported its first on 20th March 2020. Measures announced by the two countries were similar in many ways. South Africa imposed a 21-day complete lockdown on the 26th, March 2020 which was extended by a further two weeks on 9th of April. Zimbabwe, likewise, imposed a 21-day lockdown on 30th March, which it also extended by a further two weeks from 20th April. Thereafter, the countries continued with various other measures, although the periods fall outside the scope of the current study. The measures were so similar that in many quarters, Zimbabwean President Emmerson Mnangagwa was mocked for ‘copy-pasting’ the policies of his South African counterpart, Cyril Ramaphosa. Social media were awash with such comparisons, which were not common in the legacy media. The similarity in policies could have been due to the long-standing economic, political, and social relations between South Africa and Zimbabwe. Since the two countries are immediate neighbours, any laxity in control measures by one of the countries would have repercussions on the other.

This work adopted an interdisciplinary triangulated approach combining computer sciences and cultural studies world views. Computer programs written python are used to mine, clean, and analyse the vast amount of data extracted from Twitter, while at the same time the same information is subjected to a subjective and interpretive critical discourse analysis. Before the analysis is done, the LDA algorithm is used to extract topical issues discussed. Thus, the study employs a mixed-methods paradigm, in which a quasi-netnography-survey design is employed. As such, the study is epistemologically located in both media studies and computer sciences. A similar study was conducted using a related methodology, comparing newspaper coverage of the subject of third-hand smoke in China and the United States of America [1]. South Africa is chosen for the study because it is one of the countries with the most Twitter active populations [2] and the highest rate of COVID 19 infections while Zimbabwe is chosen for comparison being geographically located next to South Africa. Many Zimbabweans reside and work in South Africa [3] thus, there are many commonalities between the two countries.

2. The Media, (Mis)Information, and Health Crises

The media have always played a critical role during health crises. The mass media or legacy media set the agenda in terms of what issues citizens should be concerned with, depending on how they frame issues [4]. In this work, the authors note that while health communicators try to put across information needed to manage pandemics (for example washing hands and minimising physical contact), the mass media tend to focus on technical matters leading to the conclusion that “the media fails health services.” In other words, the mass media focus on events rather than issues [5]. At the same time, during pandemics, communication by the state and those affiliated to it becomes a form of social control as health issues matters are taken as national security concerns[6]. As a result, communication, even if through the mass media, is top-down and often paternalistic. The state’s focus is on creating and disseminating uniform messages and making sure that official comments and statistics are not questioned or critiqued. The state seeks to build consensus among citizens so that resources to fight the common threat can be mobilised without resistance. What makes social media interesting is its ability to foster bottom-up communication in which citizens proffer alternative pieces of knowledge questioning (and sometimes opposed to) the officialdom.

While the studies mentioned above go a long way towards showing the structural deficiencies of the media in communicating health information, they do not discuss alternative modes of communication, including social media. Social media have largely been hailed for their liberative potentials when compared to the legacy media, particularly in contexts where media freedoms are restrained [7]. However, the uninhibited and de-professionalised nature of social media could also breed misinformation or ‘bad advice’, that is information that “changes human behaviour to be riskier” [8]. Even when there is widespread news from official sources on measures to avoid contracting disease, bad advice can make people disregard such information and not comply with

measures taken by public health authorities. On social media, such bad advice may come in the form of memes, jokes, and forwarded (unauthenticated) and unsolicited messages that either make fun of the official sources of information, or conspiracy theories against mitigatory measures such as state-enforced quarantines, or efforts to develop vaccines. The speed and frequency at which social media produce and circulate information, including fake news, make them a phenomenon worthy of study. Previous studies have assessed these issues in the context of disease outbreaks such as norovirus, severe acute respiratory syndrome (SARS) and monkeypox [5, 6], but none have as yet focused on COVID-19, primarily because of its novelty, hence the necessity of the current study.

3. Theoretical Framework: Social Media as a Digital Public Sphere

The study is informed by the digital public sphere theory. The public sphere can be defined as an imagined space between the state and civil society where rational and diverse debates occur among citizens on issues of common interest. It is an arena of discursive interaction, sometimes critical of the state [9][10][11]. However, as Nancy Fraser [10] shows in her critique of Habermas' [9] bourgeois public sphere theory, there are many competing and intertwined forces within this arena. Some groups collude with the state in the equation of social welfare democracy. In contrast, others such as women, the peasantry and working class are marginalised that the public sphere does not represent a constellation of public opinion but only the interests of certain population groups or influential, private individuals. Fraser postulates that these competing groups give rise to counter-public sphere(s). Such diversity can be found on internet-driven social media, which counters the public sphere generated by the mass media [12]. Although the mass media also espouse ideals of a public sphere, they have become compromised by powerful corporate interests that manifest in the form of advertisements, publicity, and other commercial messages, including the top-down, uniform messages disseminated by public officials in times of crises such as the COVID-19 pandemic. Social media can be an ideal place to assess public opinion, although the media do not focus on the platforms and the topics or discourses are emanating from them. Social media platforms were essential tools in galvanising citizens against the repressive rule of former Zimbabwean President Robert Mugabe. The harsh reaction by the Mugabe regime to Twitter-organised movements such as #Tajamuka and #ThisFlag showed the importance of social media in state-civil society engagements (see Hove and Chenzi 2020). In South Africa, Twitter took centre-stage in social media movements-cum public revolts such as #Feesmustfall and #Rhodesmustfall. It ushered in a new discourse of 'fallism' that targeted minority interests and that was used as a galvanising and mobilizing force. Twitter has also been used to 'fight' cyberwars between opposing political groups [13][7]. These examples, cited together with the popular Arab Spring, show that social media constitute an emerging platform of political communication in Africa [14] which also has far-reaching implications on the field of health communication.

Studying social media should not be seen as merely examining a technology but analysing society and the unique forms and shapes that the society takes when technologized. According to Zizi Papacharissi [15], the internet revitalizes the public sphere in the form of a virtual sphere that does not manifest physically. She argues that the internet and the emergent online media create a public space, not necessarily a public sphere. Gender, race, and economic inequalities in the offline public sphere also appear in the virtual sphere in terms of access to the internet. What is pertinent in [15]'s theorisation is the ability fostered by online media for private individuals and groups to challenge the public agenda. In the context of the current study, this extends to the ability of individual netizens and groups of common interest to challenge official discourses such as official statistics of COVID-19 mortality and illnesses. It also includes information on how individuals should protect themselves from contracting the disease, as well as lockdown and self-isolation measures adopted by most governments worldwide.

Social media allow anonymity, which might foster uninhibited public opinion [15]. However, this strength is also its weakness. The people that hide behind anonymity can be termed as ‘cyberghosts’ [13]. Cyberghosts tend to trivialise rational debates by contributing vulgar and other aberrant messages. In their study of Twitter use in academia, Tomaselli and Sundar [2] note that anonymity can cause moral panic because the users often tweet without thinking. They also argue that the brevity and the influence of such ‘twirky’ (Twitter + quirky) culture that it fosters in newspapers’ letters to the editor does not enhance the public sphere. “Twirky-talk is instant, erratic and truncated. Via it, people can communicate what is simplistic, but the medium prevents deeper analysis because it denies semantic depth” [2]. Putting its limitations aside, social media is an ideal space to gauge public opinion and sentiment as this study attempts to do, in the context of the COVID-19 pandemic. While there is a scope for examining individual tweets and their embedded meanings, for feasibility reasons, the current study is delimited to topics or themes emerging from numerous computer-extracted tweets.

4. Methodology

The study utilizes a mixed-methods topic modelling approach where Python computer software was used to detect word and phrase patterns in tweets mined from the ‘South African’ and ‘Zimbabwean’ online community. The authors use the national distinctions with caution, acknowledging that even if tweets are posted from a Zimbabwean or South African internet protocol (IP) address, that does not render the tweets South African or Zimbabwean for reasons of anonymity already discussed above. The study is partly quantitative in its endeavour to find out the frequency at which certain topics/words/messages appeared on Twitter and whether the posts are original tweets or re-tweets. The appearance of certain subjects on COVID-19 at regular intervals may be assumed as a bigger concern or more attention on the issue. Thus, the regular appearance is a useful indicator of public opinion. The qualitative aspects of the study focus on the nature of discourse shared on Twitter, trying to unpack the meanings and contextualize the words or topics most tweeted about. The findings of the study are presented both statistically and narratively. Computer programs are used to mine, clean, and analyse the vast amount of data extracted from Twitter. The study can also be classified as quasi-ethnographic because it is conducted online. Netnography is “ethnographic research that combines archival and online communications work, participation and observation, with new forms of digital and network data collection, analysis and research” [16]. It uses online communication as a source of data to understand cultural or communal phenomena [17]. The complexity of studying social media lies in what [17] terms contextual sociality (consociality) where people ‘belong’ to particular communities only because they are bound by some context (what they share), which is different from the way they would belong to an ascribed identity such as race, religion and ethnicity. For that reason and the issue of anonymity, this study uses the national distinctions with caution, acknowledging that even if tweets are posted from a Zimbabwean or South African internet protocol (IP) address, that does not render the tweets ‘South African’ or ‘Zimbabwean’ for reasons of anonymity already discussed above. This makes interpretation problematic because tweets cannot be ascribed to people of a particular gender, race, ethnicity, or income group in the absence of knowledge about the ‘actual’ people that tweeted them.

Since the “power to connect” is the currency of social media belonging, the digital participation divide is a reality. While the proliferation of online media and the scramble to belong to them lead to what Kozinets terms an ‘online human’, it must be noted that many people, particularly those in the developing world, still do not have access to the internet. Thus, a constellation of opinions on social media cannot be taken as consensus because various opinions are excluded from the platform, for multiple reasons.

Tweets harvested online were treated as ‘found materials’ and were archived in comma-separated value (CSV) files. While the mined data reveal important patterns, they do not reveal the contexts in which certain words or phrases were used and how this connects to broader social, political, and economic issues. For that reason, the authors’ (subjective) knowledge based on living through the pandemic and participating on Twitter complements the big data analytics. Thus, the

study can be described as data-driven Critical discourse analysis (CDA) that it further analyses the topics and words found in the data. The COVID-19 case and the large volumes of social media data it yields present opportunities for the application of data science methodology to the study of social phenomena, as this article demonstrates. Therefore, the paper outlines an elaborate methodology, showing how data was sampled and collected, as well as the myriad of challenges implied in this approach.

4.1. Data Acquisition, Sampling, and Preprocessing

In our quasi-experiments, the paper made use of 68 000 randomly filtered messages from Twitter using the Konstanz Information MinEr (KNIME) Twitter application program interface (API), which is basically a protocol for mining data from social media platform Twitter. This was used together with a program written in Python, a popular computer programming language used to handle big data and perform complex mathematical tasks. Tweets were filtered using common hashtags (#) used during the Coronavirus outbreak, including those relating to the origins, spread and the world's response to the pandemic. The work used hashtags #COVID19SA, #CoronaVirusSA, #CoronaVirusZW, #Covid19Zim to extract initial data for both Zimbabwe and South Africa. Twitter messages were mined between March 15 and 30 April 2020. During this time, the virus had spread to all continents and killed more than 200 000 people at the time of writing. To ensure that our data was exhaustive and accurate in presenting the discussed topics, the authors further obtained related hashtags from the initially mined data. The new hashtags were subsequently used to mine more data, and the process was repeated with new hashtags obtained after each iteration until a comprehensive collection of tweets was gathered. The process, however, resulted in duplicate tweets which were later identified and eliminated before analysis.

Pre-processing was done to remove 'noise' and make the data readable by computer for easy and accurate analysis by the algorithm described above. In the context of extracted Twitter data, this work considered the following as noise: non-standard text and symbols (for example ð^w; z^w; z; &), Uniform Resource Locators (URLs, commonly referred to as web addresses) and whitespaces, hashtags, names of users and duplicate tweets. After the noise was removed, the following further pre-processing steps were conducted:

4.1.1. Lower Casing

All words were converted to lower case to reduce dimensionality and increase topic coherence. In natural language processing, words such as "Make" and "make" mean the same but when not converted to lower case, they are treated as two different words and are represented differently in the vector space used by the algorithm for analysis. However, converting everything to lowercase also changes the meaning of words or abbreviations such as UN (United Nations) or WHO (World Health Organisation). When presented in lower case UN and WHO have different meanings from the prefix "un" and the pronoun "who". To guard against possible misinterpretation, a corpus of all uppercase words was constructed before a rule-based algorithm was used to select which of the terms required changing to lowercase. The algorithm allowed us to customize its behaviour based on the output from the data.

4.1.2. Stop Words Removal

Commonly used words that do not change the meaning of the data were also eliminated from the corpus. In essence, by removing such words from the text, the topic extraction algorithm could only take as input, the most important words hence producing more accurate results. For example, given a sentence such as "What is lockdown?" The most important word is "lockdown". Given a search query for this phrase, one would want the system to focus on texts about "lockdown" over those focusing on "What is".

4.1.3. Text Normalization

Mined data contained a lot of misspellings and out-of-vocabulary (OOV) words. These included many non-English words such as people's names or nicknames written in indigenous African languages (for example Zoro, for Zororo Makamba, Zimbabwe's first COVID-19 mortality). Before normalisation, a set of all such words were extracted and identified the ones that required no standardisation. A rule-based method was then applied to normalise the remainder of the text. Text normalization was in the past effective for analyzing clinical documents where medical personnel take notes in non-standard ways [18].

4.1.4. Stemming

This reduces inflection in words (for example troubled, troubles) to their root forms (for example trouble). Word-stemming is useful for dealing with sparsity, which is a big issue when dealing with English text. Furthermore, stemming also helps with standardizing vocabulary. Twitter users may use different words and expressions to put across the same sentiment. For analysis purposes, words with different variation should be normalised and interpreted to mean the same. For instance, "drawing", "draw", "draws", "drew", "drawings" should be normalised to give the same meaning: "draw".

4.2. Experimental Setup

This section presents the experimental setup used in the work.

4.2.1. Preprocessing

Using a program written in Python, the pre-processed tweets were converted into a corpus of text, and a word cloud was used to verify that there was no unwanted data within the corpus [19]. A word cloud is a visual representation of text data, typically used to depict keyword metadata [20]. The size of each word in the illustration indicates its frequency or importance. The more important a specific word appears in text, the bigger and bolder it appears in a word cloud. A word may not be frequent yet very important due to its semantic relationship with other terms. In politics, word clouds can help analyse opinions. For instance, if "too corrupt" cropped up as a major emphasis word in campaign feedback, that should ring a warning bell. In the current study, the word cloud also served as a way of verifying the output from the pre-processing stage. Bag of words feature representation was used to represent the corpus. In this feature representation model, the corpus of all the tweets was represented as a collection of the words, disregarding grammar and word order but keeping multiplicity.

4.2.2. Topic Extraction

The Latent Dirichlet Allocation (LDA) topic modelling technique was implemented on the bag of words created from a text corpus. LDA classifies tweets into topics by building a topic per tweet and words per topic model, modelled as Dirichlet distributions. Each pre-processed tweet is modelled as a multinomial distribution of key words. The assumption is that the bag of words used to feed the model contains words that are related. In our dataset, common hashtags were used to mine the tweets; hence this assumption is valid. LDA further assumes that the data used in the analysis are made up of a mixture of topics, and the topics then generate words based on their probability distribution as shown in Fig 1. To address this assumption, tweets were mined using different, but related hashtags, each addressing a specific aspect of trending issues on the COVID-19 pandemic. Because of its ability to find hidden meaning in text, LDA has been used for topic extraction in various fields such as software engineering [21], political science [22], medicine [23] and linguistic science. In our study, LDA was configured to extract nine topics from the data for Zimbabwe and South Africa respectively. Below is a presentation and discussion of the study findings. The discussion is organised around visualisations obtained from data, namely word clouds, most common words and trigrams as well as topics extracted.

disease' before spreading to other provinces. The journalist and socialite Zororo Makamba, was Zimbabwe's first COVID-19 fatality. Makamba's celebrity status, coupled with the fact that he was well-connected among the country's social and political elites, made him a subject of many tweets and retweets. As such, it was common that whenever the disease was mentioned in those early days, a reference would also be made to Makamba. Although the microchip conspiracy theory was also prevailing in Zimbabwe, it was not frequently tweeted at the time enough to have a visual presentation on the word cloud. From both word clouds, little or no mention of nurses and doctors involved on the frontline. This might imply that people were more concerned with their individual safety than for the next person.

5.2. The Most Common Phrases

At the height of the pandemic, both President Cyril Ramaphosa (South Africa) and President Emmerson Mnangagwa (Zimbabwe) addressed their nations frequently, and citizens always looked forward to these addresses. Ramaphosa features in more than 300 tweets. This shows that people were meticulously tracking the spread of the disease. It also shows that citizens and indeed, state-sponsored or affiliated actors masquerading as private individuals converse a lot about political power. In fact, most of the talk stems from and is about those that possess political power including their decision-making, policies, and personalities.

"Vodafone boss blows" and "boss blows whistle" were also common phrases in the South African context as seen in Figure 3. These were related to the conspiracy theory that links fifth generation (5G) network development to COVID-19. The theory suggested that the 5G infrastructure being installed around the world emitted radio waves that prompted auto-immune responses which manifested as symptoms of COVID-19 in human bodies. Jonathan James, who claimed to be a former manager at telecommunications organisation Vodafone appeared in a Youtube video making the links between COVID-19 and 5G. This led to or complemented public fury over 5G, with incidents of arson and vandalization of telecommunications infrastructure being reported [25]. Assuming that this common phrase was associated with tweets by 'ordinary people', it can be argued that there is a noticeable relationship between public sentiments online and how people behave in real life. It shows the importance of big data analytics in judging and predicting public opinion as already discussed above. This also challenges Nancy Fraser's postulation that the realm of the public sphere has some "weak publics" "whose deliberative practice consists exclusively in opinion-formation and does not also encompass decision-making" [10]. In this case, the subsequent actions of arson associated with the social media outcry over 5G show a strong link between opinion formation and decision-making (revolting). "COVID19 Vaccine Africa" was another common phrase that emerged in the context of vaccine-related conspiracy theories as discussed in the South African word cloud above.

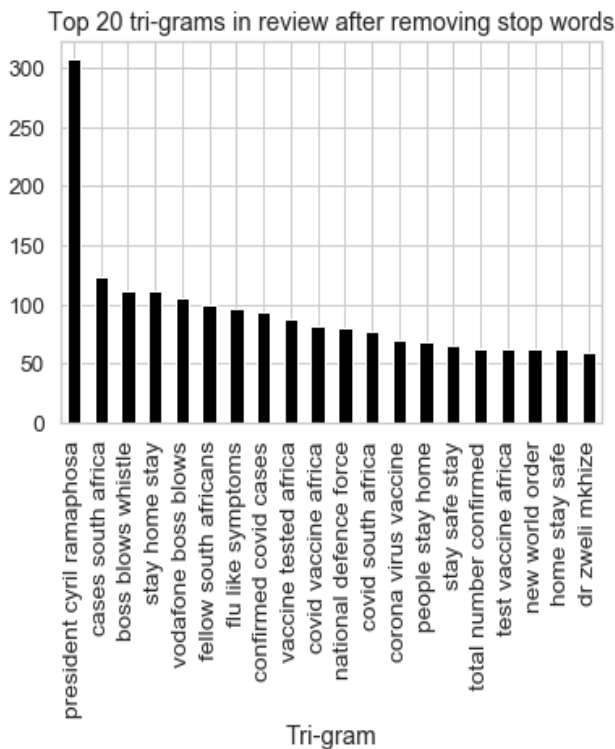


Figure 3. Tri-gram Illustrations for South African Tweets

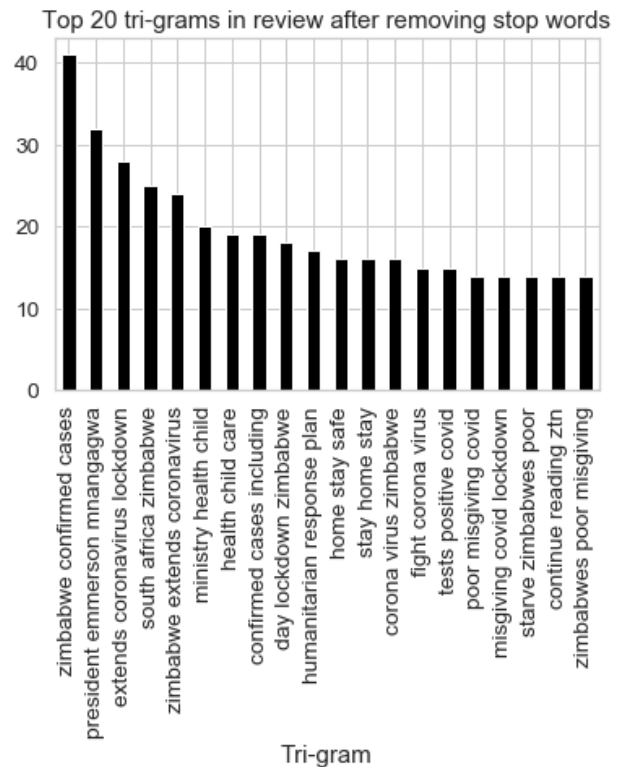


Figure 4. Tri-gram Illustrations for Zimbabwean Tweets

The most common phrase in the Zimbabwe data-set was “Zimbabwe confirmed cases”. This was in the context of doubt and mistrust concerning the official statistics issued daily by the Ministry of Health and Child Care (MOHCC). The Coronavirus update template was used by the Ministry, which Twitter users retweeted quite often uses the phrase “confirmed cases” at least twice each time. The phrase featured in more than 40 tweets. The fewer tweet counts for words and common phrases used in Zimbabwe compared to South Africa might be an indicator of internet access in the two countries, which may also translate to the frequency of tweeting. There was always a debate around the ‘confirmed cases’ with many observers accusing the government of hiding some cases that citizens had ‘confirmed’ in their communities. The word ‘confirm’ was associated with power and agency to pronounce COVID-19 statistics. When the state ‘confirmed’ cases, it meant that whatever anyone else pronounced was hearsay and could not be relied on. It is an affirmation of the paternalistic nature of health communication, which netizens contested via Twitter and other social media.

The phrase “President Emmerson Mnangagwa” was the second most common, indicating the centrality of the presidency in issues concerning COVID-19, just as was the case for South Africa. This was because it was the presidents who pronounced any national measures taken in response to COVID-19. Soon after pronouncements, the Twitter community would start dissecting the implications. The phrase was closely related to the third most popular “extends coronavirus lockdown”, indicating that the president’s extension of the initial 21-day lockdown period by a fortnight and yet another two weeks thereafter, was a subject of conjecture. The three most common phrase shows a concern with officialdom, such as cases of COVID-19 ‘confirmed’ by authorities, as well as President Mnangagwa’s authority to extend the national lockdown. This suggests that during times of crisis, there is an increase in public scrutiny of government policy and concurs with [10] assertion that “no topic should be ruled off-limits” in terms of contestation.

As seen in Figure 4 above, the fourth most common phrase “South Africa Zimbabwe”, considered in the context of the first three, indicates the concern by Zimbabweans that their country’s

response measures to COVID-19 were a copy of South Africa's. In many instances, Zimbabwean humorously referred to Mnangagwa as 'Ramakopa'. The word takes the first letters of Ramaphosa and adds 'kopa', a Shona pronunciation of 'copy', which suggested that the Zimbabwean president was a copycat. Beyond that, it mirrors the numerous commonalities between the neighbouring countries. Thus, Zimbabweans would be concerned with whatever happened in South Africa. Many Zimbabweans import essential commodities for personal consumption or resale from South Africa, thus they had to monitor developments across the border with keen interest. As the result, there was a commonplace joke that Ramaphosa should make COVID-19 related responses on behalf of both Zimbabwe and South Africa. More importantly, official statistics showed that most of the Zimbabweans who tested positive for COVID-19 were quarantined returnees from South Africa.

5.3. Analysis of the Extracted Topics

Topic extraction, a subfield of natural language processing (NLP), enables computers or machines to read, understand and derive meaning from text. NLP has been used for various tasks such as text summarization and sentiment analysis. In text-summarization computer programs are able to extract the key concepts and summarize the text just like humans would. Sentiment analysis enables computer systems to detect human feelings and emotions. Sentiment analysis has been applied in the analysis of customer reviews and political sentiments on social media. In businesses such as hotels, sentiment analysis can be crucial for the automatically handling customer feedback and promptly responding to issues that may ruin the reputation for a business [26]. Word-of-mouth rate has been found to have a strong influence on hotel occupancy hence addressing automatically detecting customer sentiments is crucial [27]. The topics presented in the tables below are derived from our own analysis of discourse represented by the common words. The computer program simply generates words per topic, although they do not name the topic. What is derived is the inference of the words based on the understanding of the context within which they were used in the two countries as seen in Table 1 and Table 2.

Table 1. Extracted Topics from South African Tweets

Topic	Common words
Topic 1: Lockdown	Lockdown; work; food; income
Topic 2: 5G conspiracy theories	Radiation; people; coronavirus; virus; towers; think; don't; cause; corona; like
Topic 3: Staying home	Home; day; stay; lockdown; safe; days; time; today
Topic 4: Implications of lockdown	Sandf; president; nothing; listen; alcohol; ramaphosa; news; remember; hands; die
Topic 5: SANDF and police violence against blacks	Sandf; saps; people; police; beat; lockdown; law; black; distancing social
Topic 6: tracing of daily statistics	Cases; workout; home; new; number; confirmed; time; covid; today; total
Topic 7: President's address and citizens' concerns	South; Africa; covid; covid_sa; coronavirus; president; africans; African; ramaphosa
Topic 8: essential workers	Lockdown; food; people; essential; workers; time; health; need; government; services
Topic 9: Bill gates and conspiracy theories	Vaccine; people; virus; don't; gates; us; bill; want; like; test

Table 2. Extracted Tweets from Zimbabwean Tweets

Topic	Common words
Topic 1: Corona virus	Zimbabwe; covid; coronavirus; people; lockdown; corona; fight; news; pandemic; country
Topic 2: Implications of	Food; people; Zimbabwe; extension; data; hunger; guys; response; build;

lockdown	week
Topic 3: Essential services	Lockdown; going; Zimbabwe; rights; human; health; covid; govt; services; year
Topic 4: Tracing of daily statistics	Zimbabwe; cases; lockdown; covid; positive; corona; coronavirus; confirmed; one; days
Topic 5: Citizens' concerns	Zimbabwe; people; us; covid; stay; social; help; need; makamba; safe
Topic 6: Staying home	Zimbabwe; lockdown; coronavirus; day; home; covid; stay; via; zororo; happy
Topic 7: Wilkins hospital	Zororo; makamba; hospital; wilkins; government; covid_19; make; name; im + dont
Topic 8: Families coping with lockdown	Zimbabwe; covid; thank; children; love; see; produce; around; made; lockdown
Topic 9: Social and economic issues	Use; households; mines; families; vulnerable; forex; group; strategic; today's; promises

The word 'lockdown' was the most common among words used in tweets related to COVID-19. For Zimbabwe, the most common word was Zimbabwe, followed by 'COVID' and 'lockdown'. This could be because the lockdown affected people's livelihoods. As a result of the lockdown, people's movements were restricted. Alcohol sales and consumption were banned in South Africa while in Zimbabwe, the initial ban was lifted after an outcry. The lockdown in some cases prevented people from moving out of their homes to purchase foodstuffs and to meet members of their social circles.

Linked to the word-cloud, the issue of conspiracy theories is 'Radiation' (from 5G) which was mentioned over 6000 times. The theory suggested that COVID-19 was not a viral illness but the body's response to radiation caused by fifth-generation technology to which many countries were implementing to improve the speed and efficiency of their telecommunication systems.

The South African Defense Forces (SANDF) and Police were also discussed in more than 4000 tweets. The defense forces and police appeared in the tweets because they were the enforcers of national lockdowns. In some instances, they were reported to have responded heavy-handedly to incidences of people disobeying lockdown regulations. At the same time, it speaks about these repressive apparatuses might have been a way of challenging state authority. The government's deployment of security personnel was met with mixed feelings.

The data shows that the word Zimbabwe features in almost all tweets. This indicates a 'national' concern in all the tweets. 'Zimbabwean' Twitter users (twimbos) were concerned about issues that affected them as Zimbabwean citizens. The word 'lockdown' also appeared in seven of the nine topics on Zimbabwe, indicating the widespread concern with this measure and giving the sense that the lockdown was always reminiscent of COVID-19. The word was accompanied by words such as 'home' 'stay' which were a common feature of the 'stay home' theme of most health awareness campaigns urging people to stay at their homes to avoid spreading or contracting the disease.

The words 'positive', 'confirmed' and 'test' also showed the concern among twimbos on the numbers of confirmed cases that had tested positive. This arose a context of mistrust with official statistics with many people on social media suspecting that the Zimbabwean government was downplaying the figures. Previously, the word 'positive' had gained popularity in describing those living with HIV. Civic society activists had managed to give the word a positive spin, suggesting that those who were HIV-positive and those around them, should not harbour negative thoughts about the condition. As such, the word 'living positively' had become popular to mean that a person who had tested positive for the virus, was living with it, but with a positive outlook. The usage of the word 'positive' in the context of COVID-19, however, had a morbid connotation to it. Due to the growing mortality rates in other countries, a positive diagnosis for COVID-19 could not be spun positively, as it meant that the patients would be put under compulsory quarantine so that they did not spread the virus.

Words such as 'food' and 'hunger' are evidence of the fears by citizens that the lockdown deprived them of freedom to source food for their families. For the Zimbabwean population, most of which is employed in the informal sector, the lockdown meant that they could not venture out to their

various income-generating enterprises. As a result, their fears of dying from hunger were real. The COVID-19 crisis presented a paradoxical situation where people could end up dying from hunger as they stayed home, trying to avoid dying from the virus.

Zororo Makamba also featured prominently in the tweets because he was the first patient to die from COVID-19. His death generated a lot of news, primarily because of his status, being a socialite, journalist and the son of a wealthy politician, James Makamba.

6. Conclusion

This work comparatively explored the variety of topics that occupied Twitter communities from Zimbabwe and South Africa. Furthermore, the work examined whether the national identities that define and differentiate citizens of these countries also exist on Twitter as evident in the emerging topics. The Latent Dirichlet Allocation (LDA) algorithm was applied to Twitter data to extract topics discussed by netizens from Zimbabwe and South Africa. The study is anchored in the digital public sphere theory, which treats social media applications as virtual platforms where netizens commune to share ideas and debate about issues that affect them. From the collected data, it is evident that there are common fears across the two African nations. The two countries do not only share a border but the same culture. The findings from this work reflect the strong relationship between these two nations. Furthermore, the novelty and fatality of COVID-19 have bred fear amongst the citizens from both countries. Consequently, there is evidence of social media being used for the propagation of doomsday conspiracy theories. Another important observation is that citizens were meticulously tracking the spread of the disease. Results from this work are crucial in informing decision makers in policy formulation and designing response mechanisms to the COVID-19 pandemic. Analyzing how citizens respond to events is crucial for government as it informs how citizens are likely to react to policies.

References

- [1] Q. Liu, Q. Chen, J. Shen, H. Wu, Y. Sun, and W.-K. Ming, "Data analysis and visualization of newspaper articles on thirdhand smoke: a topic modeling approach," *JMIR medical informatics*, vol. 7, no. 1, p. e12414, 2019.
- [2] K. Tomaselli and T. Sundar, "Twitter and African Academia," *Editors' Bulletin*, vol. 7, no. 3, pp. 101-104, 2011.
- [3] N. De Jager and C. Musuva, "The influx of Zimbabweans into South Africa: a crisis of governance that spills over," *Africa Review*, vol. 8, no. 1, pp. 15-30, 2016.
- [4] J. Leask, C. Hooker, and C. King, "Media coverage of health issues and how to work more effectively with journalists: a qualitative study," *BMC public health*, vol. 10, no. 1, p. 535, 2010.
- [5] T.-J. Shih, R. Wijaya, and D. Brossard, "Media coverage of public health epidemics: Linking framing and issue attention cycle toward an integrated theory of print news coverage of epidemics," *Mass Communication & Society*, vol. 11, no. 2, pp. 141-160, 2008.
- [6] I. Basnyat and S. T. Lee, "Framing of Influenza A (H1N1) pandemic in a Singaporean newspaper," *Health promotion international*, vol. 30, no. 4, pp. 942-953, 2015.
- [7] A. Chibwe and O. Ureke, "'Political gladiators' on Facebook in Zimbabwe: a discursive analysis of intra-ZANU-PF cyber wars; Baba Jukwa versus Amai Jukwa," 2017.
- [8] J. Brainard and P. R. Hunter, "Misinformation making a disease outbreak worse: outcomes compared for influenza, monkeypox, and norovirus," *Simulation*, vol. 96, no. 4, pp. 365-374, 2020.
- [9] J. Habermas and J. Habermas, *The structural transformation of the public sphere: An inquiry into a category of bourgeois society*. MIT press, 1991.

- [10] N. Fraser, "Rethinking the public sphere: A contribution to the critique of actually existing democracy," *Social text*, no. 25/26, pp. 56-80, 1990.
- [11] S. Splichal, "Eclipse of „the Public”,“ *The Digital Public Sphere*. Goteborg: Nordicom, pp. 23-40, 2010.
- [12] V. Sampedro and M. M. Avidad, "The digital public sphere: An alternative and counterhegemonic space? The case of Spain," *International journal of communication*, vol. 12, p. 22, 2018.
- [13] A. Chibuwe, "Social Media and Elections in Zimbabwe: Twitter War between Pro-ZANU-PF and Pro-MDC-A Netizens," *Communicatio*, pp. 1-24, 2020.
- [14] H. Wasserman, "Media, conflict and democratisation in Africa: Political communication by other means," 2019.
- [15] Z. Papacharissi, "The virtual sphere 2.0: The Internet, the public sphere, and beyond," in *Routledge handbook of Internet politics*: Routledge, 2008, pp. 246-261.
- [16] R. V. Kozinets, *Netnography: The essential guide to qualitative social media research*. SAGE Publications Limited, 2019.
- [17] R. V. Kozinets, *Netnography: Doing ethnographic research online*. Sage publications, 2010.
- [18] A. Mykowiecka, M. Marciniak, and A. Kupść, "Rule-based information extraction from patients' clinical data," *Journal of biomedical informatics*, vol. 42, no. 5, pp. 923-936, 2009.
- [19] M. Sponder and G. F. Khan, *Digital analytics for marketing*. Routledge, 2017.
- [20] F. Heimerl, S. Lohmann, S. Lange, and T. Ertl, "Word cloud explorer: Text analytics based on word clouds," in *2014 47th Hawaii International Conference on System Sciences, 2014: IEEE*, pp. 1833-1842.
- [21] A. Hamdy and M. Elsayed, "Topic modelling for automatic selection of software design patterns," in *Proceedings of the International Conference on Geoinformatics and Data Analysis, 2018*, pp. 41-46.
- [22] Y. Fang, L. Si, N. Somasundaram, and Z. Yu, "Mining contrastive opinions on political texts using cross-perspective topic model," in *Proceedings of the fifth ACM international conference on Web search and data mining, 2012*, pp. 63-72.
- [23] R. Cohen, I. Aviram, M. Elhadad, and N. Elhadad, "Redundancy-aware topic modeling for patient record notes," *PloS one*, vol. 9, no. 2, p. e87555, 2014.
- [24] M. Okwonga, "The Frency doctors who wanted to test vaccines on Africans and Western medicine's dark history' in Quartz Africa." [Online]. Available: <https://qz.com/africa/1836272/french-doctors-say-test-covid-19-vaccine-on-africans-spark-fury/>.
- [25] T. Tahir, "Flaming Idiot: 5G Coronavirus Conspiracy Theorist who Claimed to be Vodafone Boss is Revealed as a Preacher from Luton'," 2020. [Online]. Available: <https://www.thesun.co.uk/news/11476133/5g-coronavirus-conspiracy-theorist-vodafone-preacher-luton/>.
- [26] T. Mullen and R. Malouf, "A Preliminary Investigation into Sentiment Analysis of Informal Political Discourse," in *AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs, 2006*, pp. 159-162.
- [27] G. Viglia, R. Minazzi, and D. Buhalis, "The influence of e-word-of-mouth on hotel occupancy rate," *International Journal of Contemporary Hospitality Management*, 2016.