Tackling the Spread of Misinformation on Twitter: Topic modelling and Information Retrieval to Match Tweets With Official Information

Ekaterina Batueva (ekba21ab) & Pernille Opstad (peop15ab)

Copenhagen Business School

Characters Maximum: 34215

Abstract

The purpose of this work is to develop an approach to informing Twitter users about the presence of false information in a tweet as well as providing reliable information scraped from the WHO website to reduce the spread of misinformation in Twitter as well as to increase users' awareness. A dataset (consists of 213 069 and 71 614 observations, before and after preprocessing, respectively) consisting of information about users and their tweets containing the words "plandemic" and "scamdemic" for the period from March to June 2021 was used for topic modeling. Two techniques were used, LDA and top2vec, to compare the quality of extracted topics on short text and to determine applicability of topics as queries to retrieve reliable information from WHO. It was concluded that topics derived using LDA model are more sufficient as queries for information retrieval from WHO data.

Link to google repository.

Keywords: Covid misinformation, #scamdemic, #plandemic, Twitter, topic modeling, top2vec, LDA, information retrieval

1. Introduction

The Covid-19 pandemic has led to the spread of "infodemic" among the population (Bruns et al., 2021), which refers to excessive false or misleading information during a disease outbreak (WHO, 2020). Conspiracy theories are actively spread in social media (Fuchs, 2021) due to increasing contagion and stickiness (Dow et al., 2021). Moreover, studies show that people trust the opinion of celebrities more than the opinion of official representatives (Honora et al., 2022). It leads to several consequences. Firstly, the difficulty of finding trustworthy sources in the flow of incorrect information (WHO, 2020). Secondly, it was discovered that belief in Covid conspiracy theories has important social consequences, such as loss of trust in government agencies (Pummerer, 2021).

To tackle the increasing volume of posts with misleading information, Twitter has created "COVID-19 misleading information policy" (Twitter), where punishment as well as informing systems are provided (Ibid.). Nevertheless, Rossi (2022) argues that Twitter failed to ban accounts that spread repeatedly conspiracies theories since only 12% of accounts were suspended.

During the past 3 years, a great variety of Covid 19 conspiracy theories was created, such as intentional creation of Covid by government of China (Douglas, 2021) or the danger of vaccines due to belief that they change human DNA (Pertwee et al., 2022). Popular hashtags for sharing covid-19 related consipracies on Twitter are "plandemic" and "scamdemic" (Ittefaq Abwao, 2021).

2. Research Questions

The goal of our project is to develop a method of informing users about the presence of covid misinformation in tweets containing "plandemic" or "scamdemic" words by placing a warning label as well as providing reliable Covid information scraped from the WHO website as a reactive mechanism on spread of misleading information in Twitter.

To achieve this goal, the following research questions should be answered:

1. What topics can be derived from Covid-19 conspiracy related tweets?
2. What topic modeling technique will provide better interpretability of topics on short and informal text?
3. To what extent does World Health Organisation cover found topics in the question-answer section on their website?

4. To what extent the found topics are suitable as queries to retrieve relevant information from World Health Organisation, and how does it compare to querying tweets directly?

3. Related Work

A spread of fake health-related information is a prominent problem and studies on health-related misinformation are aimed at learning the mechanisms and patterns of misinformation spread (Wang et al., 2019). One of the most popular research topics is recognition of misinformation and bot detection where machine learning and deep learning are used simultaneously with NLP techniques (Shahid et al., 2022). Popular methods of analysis used in this type of research are sentiment and content analysis, where NLP is used for preprocessing (Wang et al., 2019). Despite the simplicity of some methods, such as the bag of words, it is possible to achieve high accuracy in combination with machine learning (Hamid, 2020). Nevertheless, more sophisticated approaches are developed. For instance, Nigam (2021) used supervised deep learning algorithm, pre-trained BERT, to detect Covid conspiracy data using cosine similarity of texts.

Covid-19 as a multifaceted topic is discussed at the junction of many different topics, including politics and economics. Thus, more research began to appear on topic modeling of health-related discussion. Most of the research use traditional topic modeling techniques, such as LDA, to analyze the general reaction of people on Covid (Xue et al., 2020; Lanier et al., 2022). Nevertheless, new approaches and models are developed to be more suitable to short text analysis (Egger, 2022). Thus, Mackey (2020) developed an unsupervised machine learning biterm topic modeling that is suitable for clustering short texts such as tweets for further analysis. Moreover, it was shown (Paul Dredze, 2014) that statistical topic model can be used to identify topics in Twitter data with minimal human supervision.

4. Theoretical Background

4.1. Topic Modeling

Topic models discover the latent semantic structure within a collection of unstructured texts. The information contained in the documents are represented as weighted sets of words, from which topics are inferred to organise and identify relevant documents as well as provide a high-level summary of the contents (Angelov, 2020).

Latent Dirichlet Allocation (LDA) is typically used for topic modelling. The three-level hierarchical Bayesian model uses a generative probabilistic approach, in which texts are described as a set of topics and a distribution of word probabilities associated with each topic (Blei et al., 2003; Angelov, 2020). LDA interprets the topic space into a discrete space that assumes a specified number of topics is known, and uses a bag-of-words representation of texts which requires handling of stop-words as well as stemming and lemmatization techniques (Angelov, 2020).

While the goal of LDA is to be able to recreate the documents' word distributions, Top2Vec focus on the representation of semantic similarity of words, sentences, or documents within spatial proximity (Egger, 2022). This is done by generating jointly embedded word and document vectors, and clustering the vectors (Angelov, 2020). Representing similar documents as dense areas, for which topic vectors are centroids of the areas and thus the average document to represent the cluster. The closest word to topic vectors are more descriptive of the topic and cluster of texts (ibid.). Besides from identifying an unspecified number of topics, Top2Vec allows to find similar documents and words including searching for documents and topics given a keyword (Documentation). Thus, different types of insights can be generated from the two methods.

4.2. Information Retrieval

Information retrieval (IR) concerns finding relevant documents from a large collection, often unstructured texts, that is relevant to a specific information need (Manning et al., 2008). The retrieval process is based on queries of index terms, and assumes that a combination of terms can characterize the semantics of both documents and required information with the goal of evaluating relevance in terms of matches between the search terms and the documents (ibid.). Precision and recall are appropriate to measure the relevance of the documents returned by the IR system, meaning the ratio of found documents where relevant and ratio of the relevant material that was found, respectively (ibid.).

There are several types of information retrieval methods. Since Boolean method does not have a means for determining evidence or importance of results, ranked retrieval is preferred in this context. The idea of *ranked retrieval* is that a document or zone that mentions a query term more often are assigned a higher score to indicate higher importance with the query. The concept of document vectors to describe the relative importance of the terms in a document, measured by a composite of term frequency and inverse documents frequency, td-idf (ibid.). The representation of the document vectors in a common space is the basis for the *vector space model*, for which also queries are represented as vectors and allowing to measure cosine similarity as a relevance score.

5. Methodology

Figure 1 gives a high-level visualisation of the process steps in our experiment: (1) Data collection uses tools for scraping, (2) preprocessing uses tools for text cleaning, word normalisation and handling of emoticons, (3) LDA and top2vec for topic modelling and, (4) Pyterrier library for Information Retrieval.

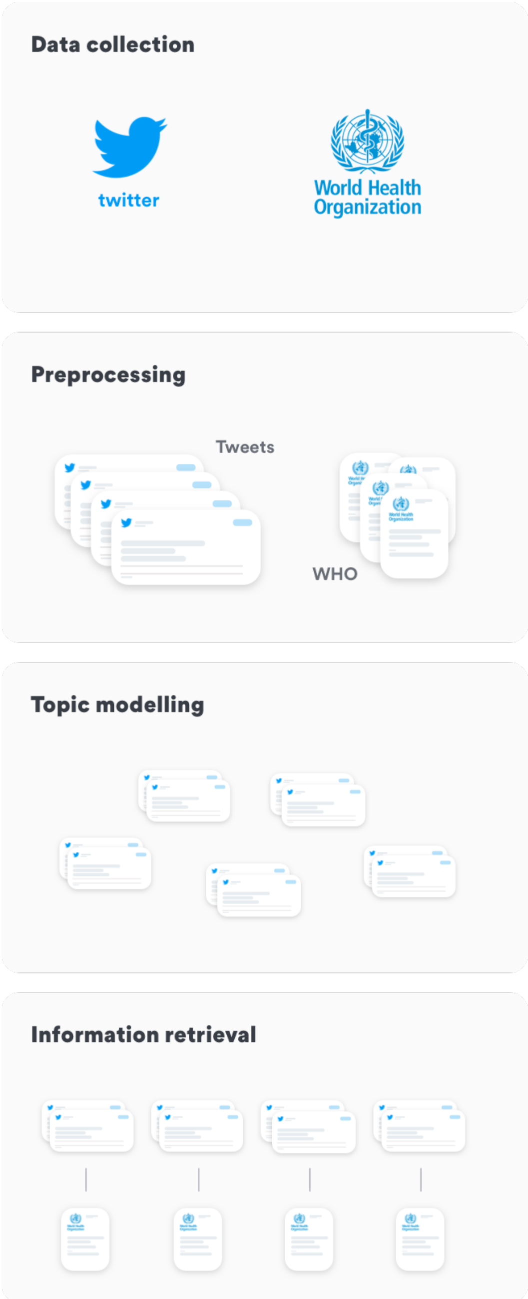


Figure 1: High-level method outline

- `twitterdata` folder contains tweets for plandemic and scamdemic hashtags separated into weeks.
- `tweet_preprocessing.py` is a module with implemented functions for preprocessing tweets, imparted and used in all remaining notebook files.
- `Exploratory-analysis.ipynb` contains analysis of the tweets.
- `LDA.ipynb` and `top2vec.ipynb` contain topic modelling, and pickle files ldamallet.pkl and t2v.pkl are saved topic models.
- `WHO-IR.ipynb` contains scraping of WHO data, preprocessing and information retrieval.

5.1. Data Collection and Description

Tweets tagged with "plandemic" and "scamdemic", collected for an ongoing research project at CBS. A subset of the data is described in Rossi (2022). The dataset consists of weekly collected tweets, which were published with popular Covid-19 conspiracies hashtags "scamdemic" and "plandemic", as well as accounts that created tweets or interacted with them during spring of 2021, namely from March to June. During this period 105 482 tweets were published using "plandemic" hashtag and 107 587 tweets contained "scamdemic" hashtag. Thus, the merged dataset has in total 213 069 instances.

Official information from World Health Organisation concerning Covid-19 is provided in the extensive Q&A section available on their website. The dataset consists of 56 main topics with a total of 521 questions and answers. The library Beautiful Soup is used for scraping information from the website, and is contained in WHO_IR.ipynb. The approach is to first iterate through the anchors elements of the div-section containing the main topics, to retrieve links for each topic's Q&A section. This code is encapsulated in function `scrape_QA_topics()` Followingly, in function `scrape_QA_answers()`, each link is visited in which each question and answer is saved into a nested dictionary.

5.2. Data Pre-Processing

Besides the length of the texts, the data sets are different in the degree of formality. While WHO-data has a high extent of proper spelling and punctuation, tweets have a relatively high degree of abbreviations, misspelling and made-up words as well as extensive use of emojis. The twitter data is thus prepared for process modelling in six high-level steps contained in the `tweet_preprocessing.py` module:

Extracting features: User-ID and post-ID is extracted from the link as they are useful for identification of the posts. The tweets contain hashtags, user tags, links and re-tweet indication. These are extracted and removed from the tweet into separate columns as they hold useful information but are not particularly valuable for the text modeling. The hashtags are included in the text without the hashtag symbol.

Handling of emojis: Informal communication on twitter and other social media platforms often include symbols and non-textual expressions. Wegrzyn-Wolska et al. (2016) proposes three methods for handling emoticons in sentiment analysis: (1) deletion (2) translating to labels and, (3) translating to textual descriptions. In this case for topic modelling emoticons and emojis are handled with the emoji library where they are translated to text description using their defined emoji dictionary. However, differing from the approach of Wegrzyn-Wolska et al. (2016) the descriptions are read as a single token because we want to discover topics or meaning behind the emojis in the particular context.

Cleaning tweets: The first cleaning step is to remove links, hashtag symbols and user names as proposed by Magliani et. al. (2016). Numbers and special characters are removed to reduce noise in tweet data. However, numbers consecutive to words are maintained because they can be relevant for the context or be read as letters or words in informal digital communication. Non-ASCII symbols are filtered as the topic modeling will be limited to English-based alphabet only and text is transformed to lower capitals.

There are some challenges of working with informal communication texts besides use of symbols, emoticons and numbers instead of letters. In particular, there are relatively more cases of misspelling of words, wrong use of spacing and special characters than found in literature, newspapers and scientific papers. Therefore, the cleaning step is iterated to achieve better tokenization.

Removing irrelevant instances: Duplicate tweets, as well as tweets shorter than 10 characters, are removed to achieve more meaningful training data and avoid overfitting.

Concentration: This step concerns shrinking the vocabulary space. Stop-words are removed because they do not hold particular information about context or topic of the tweets. The texts are lemmatized to achieve fewer variations of the same words, which is particularly present in tweets with varying levels of informality.

Tokenization: The tweets are separated into lists of words for input to LDA topic modelling.

Since the WHO-data is relatively clean, it is simply preprocessed by combining the questions and answer columns and removing escape sequences that were extracted when scraping, removing punctuation and stop-words, and performing lemmatization to reduce the variation of the same words.

5.3. Exploratory Analysis

As the first step in analyzing tweets, a word cloud was built that displays the most frequently encountered words. Figure 2 shows that the most repeated words are related to the pandemic: 'covid', 'vaccine', 'people', 'lockdown'.

Figure 2: The most used words in collected tweets

Nevertheless, 2 other categories can be traced in this word cloud, namely politics and conspiracy. Frequently used words referring to politics are 'Trump', 'election', and 'politician'. To express conspiracy ideas, users chose the following words: 'truth', 'planned', 'wake up', 'agenda' and 'hoax'.

After preprocessing the data, the hashtags were separated for further analysis. The dataset contains 13 300 unique hashtags, while 'scamdemic' and 'plandemic' is the most popular, having 19 663 and 15 439 mentions, respectively.

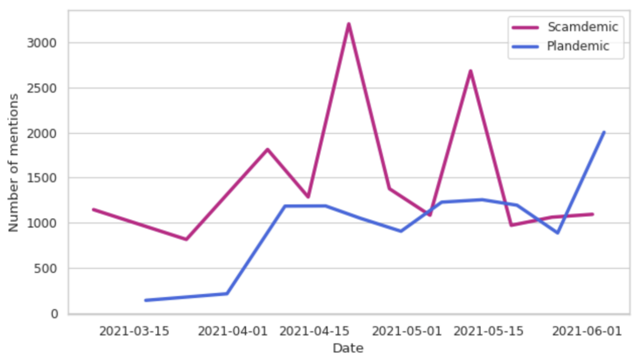


Figure 3: Weekly frequency of the use of hashtags scamdemic and plandemic

Figure 3, illustrating the dynamics of the use of these hashtags, shows that 'scamdemic' was used more often in non-repeating tweets while 'scamdemic' has several spikes associated with the reaction to news about the disease. The biggest surge in April was caused by the recommendation to suspend the use of the Johnson & Johnson vaccine because of possible blood clots (CDC, 2022). A possible reason for the more smoothed graph of the hashtag plandemic may be more strict monitoring of the use of it by Twitter's algorithms due to bigger popularity of the hashtag.

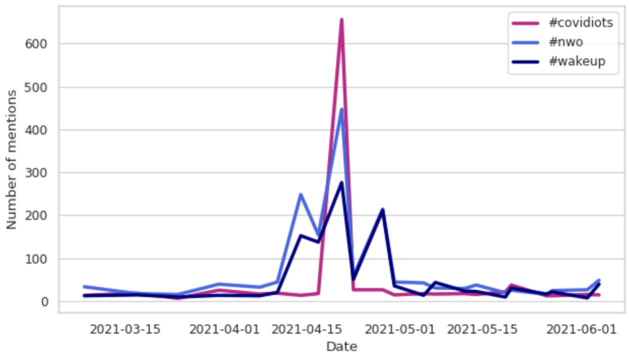


Figure 4: Weekly frequency of the use of hashtags scamdemic and plandemic

Additionally, from analysis of the 50 most popular hashtags it was found out that 26 of them relate to conspiracy. Visual analysis revealed that most conspiracy hashtags have a very short life cycle of about 1 month with one or more peaks of use that are caused by news. This pattern is shown on the chart 4. The whole list of the most popular hashtags as well as the frequency of their use in the period from March to June 2021 are presented in an appendix A.

6. Topic Modeling

6.1. LDA

To determine the optimal hyper-parameters of the LDA model, the coherence score metric was used since it allows highlighting the most informative topics that correlate well with the opinion of experts (Mimno, 2011). The model with the highest score, 46.87%, was the model that identified 12 main categories of tweets. A table with extracted topics as well as with the 20 most significant words in each topic is presented in Appendix B.

Doubts about medical recommendations: the main dissatisfaction with medical recommendations was directed to guidelines on vaccination and wearing masks. Topic 4 is accompanied not only by conspiracy terminology about vaccination, but also by calls to refuse vaccination: 'justsayno'. Twitter users also actively expressed a negative attitude towards wearing masks, which is showed by the choice of words, for example, 'bullshit' and 'sheeps' in combination with 'maskoff' in topic 6.

Mortality due to covid: The topic of mortality prevails in topic 10. It is worth noting that tweets differ significantly in the transmitted idea; some users support conspiracies about mortality due to vaccination, while another part of users reports on the tragic consequences of the virus itself.

Condemnation of state authorities: 3 topics out of 12 are directly related to the condemnation of the authorities and their actions. Topics 0 and 7 are similar emotionally, where the rhetoric is moderately negative. This similarity is also noticeable by proximity of the corresponding bubbles on the Intertopic Distance Map, which is posted in Appendix B. The topic 0 is more focused on expressing opinions about the extraordinary control by the state, while the 7th topic reveals feelings about the belief in depopulation plans and economy reset. Although topic 2 is also about the attitude of the people to power, used words are extremely negative: 'crimes against humanity', 'evil', and 'scam'.

Distrust of the authorities: the general distrust of the authorities can be traced in topics 1, 3 and 9, which are located not far from the previous topics. Topic 1 is related to the negative attitude of people to the benefits of pharmaceutical companies. The main words are 'make', 'billions', 'pharma' and 'money'. At the same time, the tweets from topic 3 are aimed at exposing states. This is evidenced by the following words: 'truth', 'theory', 'fact', 'video', and 'false'. Topic 9 is aimed not so much at discussing the current government and their actions, as at discussing politics in the United States as a whole. The confrontation of the two parties and mutual accusations are noticeable. Many expressions are offensive or have a negative tone: 'fraud', 'communist' and 'libtard'.

Emoticons: Topic 11 combines emojis used by users. Despite the fact that the highlighted emoticons are positive, they are used in a negative context - mockery and irony over others. This is also evidenced by the frequent use of swear words in these tweets.

6.2. Top2Vec

The top2vec model was defined with pre-trained embedding model "universal-sentence-encoder-multilingual" because it is faster and suggested for smaller data sets, as well as for multilingual corpora (Angelov, 2022). `min_count` is set to five to accommodate a smaller corpora (ibid.). The model found an initial 309 topics which was reduced to 12 topics after trying a handful of different number of topics. From manual inspection 10-15 topics gave the most distinct topics with less thematic overlap, while still allowing a range of different themes to be captured. A table with extracted topics as well as with the 20 most significant words in each topic is presented in Appendix C. From the found topics, displayed in figure 4 and 9, general themes and opinions can be interpreted.

Doubts about medical recommendations: As expected, masks and vaccines were prominent topics found in 4 of the 12 reduced topics, described with words of negative connotation like 'masksoff', 'fuckmasks', 'deaths' and 'kills'. Keyword search of the term 'vaccinekills' interestingly shows close relation to terms 'vaccineswork' (0.92) and 'vaccineforall' (0.89), possibly indicating occurrences of nuanced discussions or sarcasm.

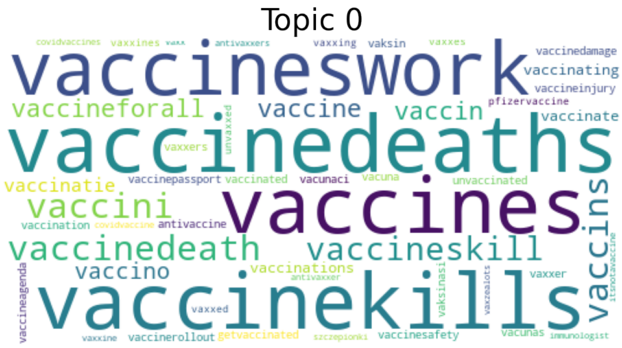


Figure 5: Vaccine hesitancy

Governmental skepticism: From topic 2 it is evident that the general political opinion is negative to democrats and their electorate win, with words like 'fraud', 'trumpwasright', and 'libtards'. Topic 9 is connected to bankruptcy and socioeconomic class with a clear negative sentiment directed towards the upper class with words like 'greed' and 'profits', while topic 8 seems to have relatively natural mention of the globalisation perspective of the pandemic with only slight association to extremist terms like 'tyranny', 'fascism' and 'terrorism'.

Skepticism about mainstream media and science: Further, topic 7 portrays a clear scepticism towards the intentions of the mainstream media and generally accepted information, referred to as 'truthers', as well as denial of generally accepted science.

Document search by keyword showed the closest related post to 'trumpwasright', "fauci busted trump was right", is related to [Anthony Stephen] Fauci and also with references to the alleged election steal. A post related to keyword 'electionfraud' referred to concerns of illegal ballot counting and secret networks with political control referred to as 'deep state'.

Rhetoric: Topic 1 and 5 expresses a degrading and negative rhetoric with prominent slander terms like 'scumbag', 'dipshit', and 'bullshit', while topic 11 have more neutral argumentative nature.

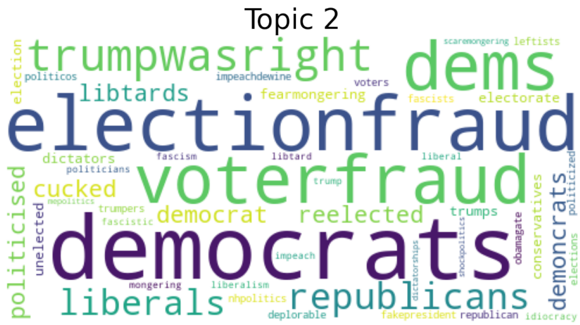


Figure 6: Political and news related topics

6.3. Comparisons of LDA and Top2Vec

We have combined all the topics into more extensive categories, displayed in table 1. It is evident that the topics had categorical similarities and yet some differences. As expected, both models highlighted topics expressing skepticism about medical guidelines and related to politics and anti-state rhetoric. LDA presents more topics within politics theme as well as a separate topic related to emoticons, while top2vec has more anti-vaccine and anti-masks topics.

Table 1: Comparison of selected categories by LDA and Top2Vec models

Category	LDA	Top2Vec
Vaccines	t4, t10	t0, t4, (t10)
Masks	t6	t6, (t10)
Politics and society	t1, t0, t2, t7, t9	t2, t8, t9
Information channels	t3	t7
Rhetoric and language	-	t1, t5, t11
Emoticons	t11	-

The main differences between these techniques can be seen in the distribution of the most significant words in topics provided in Appendix B and C. Due to the data preprocessing, including lemmatization of words, the dictionary used for constructing the LDA model consists of more generalized words. Thus, the model does not detect differences in the use of emotionally colored variants of the same word, unlike the top2vec model. Moreover, LDA's dictionary of normalized words could be more suitable for information retrieval.

Due to the poor performance of the LDA model on short texts, such as Twitter (Xu et al., 2018), derived topics 5, 8 and 11 are meaningless for the following stage of the research, despite being combined according to clear characteristics.

Overall, one of the main differences is that LDA provides topics with more general vocabulary, while top2vec topics are described by terms more characteristic for tweets.

7. Information Retrieval

To perform information retrieval the existing Pyterrier library were utilized, where the `DFIndexer()` takes the data-frame with preprocessed documents and indexes them. From the 526 documents 3 735 distinct terms are found and organised to a dictionary with 28 931 number of associated postings.

Using the generated index, the function `BatchRetrieve()` is used for simple retrieval of documents based on a given query. This is implemented in function `retrieve_QA_from_words()` to take a text string as a query and return the matching document's IDs and scores. Since we used the default parameter, the score represents the ranking score of weighing model 'BM25'. Scores can be obtained for the more traditional tf-idf weights by re-ranking. The implemented function `words_frequency_in_documents()` also takes a string and uses the properties of indexing to find each term's frequencies in documents, as well as the document IDs and number of documents they occur in, which can be applied between any indexed corpus and cleaned string.

1. Topic coverage in WHO texts
The first information retrieval experiment concerns finding documents that matches topics. More specifically, for each topic the ten most important terms are used as a query passed to the `retrieve_QA_from_words()` function, which identifies the matching document in the WHO corpora with the highest score. Further, running function `words_frequency_in_documents()` allows inspection of term matches for each topic.

Top2vec topics are matched with documents of varying relevance and upon closer inspection it seems the assigned score is not necessarily a good indication of relevance. For instance, the politically related topic (2) is also matched with a quite high score, with only one matching term 'democrat' referring to the 'democratic republic of Congo' - seemingly irrelevant for the overall context of the topic. Secondly, it seems that topics with more informal and twitter characteristic terms are more difficult to match with texts with more formal language, like 'fake-science' and 'vaxxers'. However, it seems that the more general the term, the more irrelevant or random texts are matched, for example match of topic 11 on term 'earlier'. From figure 11 it can be observed that topics concerning vaccines and epidemics terms, namely topic 0 and 4, are matched a document with relatively high score. Topics covering normal but specific terms like 'globalisation', 'epidemic', 'vaccines' and 'masks' were matched appropriately, with score ranging from 10.73-4.78.

Information retrieval on LDA topics yields document matches as shown in figure 12 and match score of topics in figure 7. From this it is evident that all topics were matched with a significantly higher score than seen in top2vec topic matches. Furthermore, all topics where matched and upon closer inspection of the topics they seem to be relevant. For example, topic 0 about vaccines and government control are matched with document about government and decision making, and topic 3 concerning distrust in information sources was matched with document about source reliability and information channels. Looking closer at the topics with lower score it is clear that a few topics does not match well with the retrieved information. For instance, topic 11 with emoticon related terms are matched with document about mask usage and topic 9 about election and politics are matched with a section about domestic violence, showing that some topics are not well covered in the WHO data or that they are not suitable for information retrieval or to catch context.

Comparing information retrieval on top2vec and LDA topics, it is evident that LDA retrieves more relevant documents and with higher coverage as seen in figure 7. This can be due to more general terms derived from LDA, while top2vec found more twitter-specific terms and informal language.

2. Document retrieval of tweets
In the second information retrieval experiment a random sample of the cleaned tweets are queried in the WHO index, for which figure 13 shows the resulting documents and scores. On a high-level inspection, it is observable that shorter tweets generate matches with lower scores while higher scores are more common for longer tweets, which is sensible as more terms increases chance of matching terms in the index.

Further, compared to matches for LDA topics, direct querying with tweets result in lower scores and seemingly less relevant results upon closer inspection of tweets and retrieved documents. Also document matches with the relatively high score can produce non-relevant results. For example, a match with score 12.96 of tweet about economic reset and government control (nr 21 in

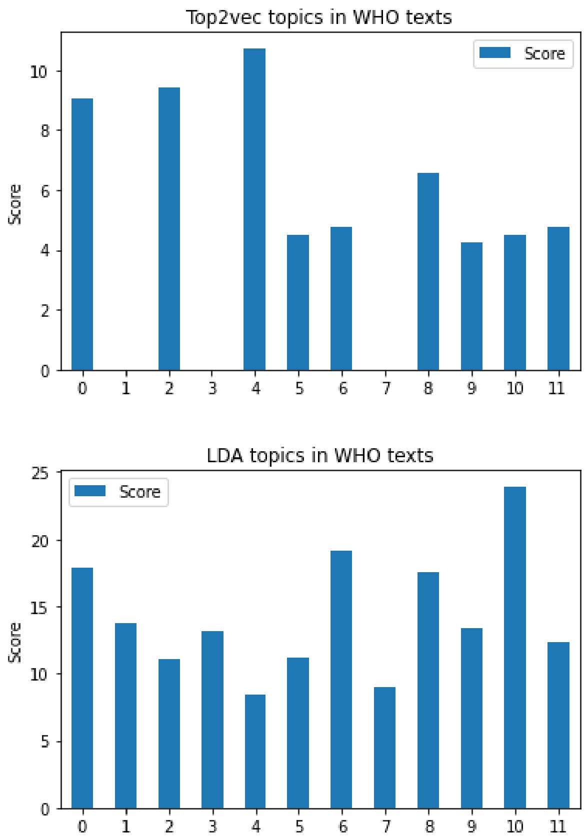


Figure 7: Document matches or top2vec and LDA topics

figure 13, is inaccurately matched with document about experiencing death among friends and family.

8. Discussion and conclusion

This project focused on creating a reactive approach to reducing the spread of Covid-related misinformation on Twitter. The provided method consists of two phases. The purpose of the first phase is to identify the discussed topics based on tweets containing the 'plandemic' and 'scamdemic' words. The second phase consists of retrieving the most relevant query containing reliable information from WHO's question-answer section.

In the first part of the project, two methods for determining topics in documents were used, namely LDA and top2vec. Table 3 and 4 shows that LDA and top2vec identified thematically similar topics related to vaccines, masks and government control with negatively associated terms. Nevertheless, in top2vec there were also topics related to information channels and socioeconomic topics.

However, the terms associated with the topics are differs substantially. It can be seen that LDA catches more general terms while top2vec catches more twitter-specific rhetoric, from which it is easier to derive an opinion or emotions. For example, from several topics emotions of distrust, fear, anger and despite were easily interpretive without knowing the context of the data. Thus, it was found that in this case top2vec better captures nuances and groups catchphrases, while LDA is better at generalizing the biased and informal language to create a better basis for information retrieval of official information sources.

Information retrieval part consist of conducting two experiments: topic coverage in WHO texts and querying WHO answers using topic terms benchmarked against querying directly with cleaned tweets. Comparing the results from top2vec and LDA when testing the 1st hypothesis, where the 20 most important terms were used to query the indexed WHO documents, it is apparent that using LDA is more sufficient as queries for information retrieval from the WHO corpora. LDA produced higher scores and higher coverage of topic matches, likely due to the topics being characterised by more general vocabulary, while still providing a specific context.

Meanwhile, top2vec produced more specific vocabulary that are characteristic for twitter language, like hashtags and composite terms. Although these terms did not result in appropriate document matches in the WHO vocabulary, it still has potential use for assessing the gap between trending misinformation and official information sources. Possible applications might involve grouping of hashtags to tag individual text segments in the Q&A section in order to improve relevance of document matches. It also creates a basis to understand areas and trends within misinformation and create content to target this specifically. Another application is to find related terms to improve twitter's algorithm to suppress or remove posts that can be categorised as misinformation. This could be explored in future research.

Finally, it was found out that important terms from found topics as queries in the WHO index produced better document matches compared to using tweets directly. This is indicated by the significantly higher score and more term matches, and is confirmed in a manual assessment of document retrieved by each topic.

It can be concluded that the presented method of using the 20 most significant words from topics, extracted by the LDA model, can be used to query relevant information from WHO data. Thus, it can be used as a measure to provide more specific information related to the respective tweet as a way to prevent the spread of misinformation and increase user awareness of false content.

9. Limitations and Future Work

During tokenization in data preprocessing part emojis in text format were divided into separate words at the connection by "_". In this regard, some words appeared in the dictionary that do not carry a semantic load, but could affect the performance of the LDA model.
Nevertheless, the experiments showed a potential method for improving the misinformation alerts with more specific information through retrieval on topics rather than specific tweets. However, it assumes that the topic modelling, in this case LDA, accurately maps tweets to meaningful topics. Furthermore, as it was shown in table 3 and 4 not all topics are meaningful to the pandemic context.

The scope of the topic modelling phase of the project is limited to the use of top2vec and Latent Dirichlet Allocation method. However, recent have acknowledged BERTopic to encode contextual information from short and unstructured text like tweets(Egger & Yu, 2022). Furthermore, LDA uses coherence score to assess semantic similarity while top2vec relies on manual inspection to find the appropriate number of reduced topics. Calculating the information gain, e.g. 'mutual information' could be useful to properly assess how well found topics by top2vec describe the twitter dataset (Angelov, 2020).

Lastly, it should be addressed that the method applied in this project is based on manual assessment of matches between query and retrieved documents are based on subjective interpretation and understanding of the topics and the context it is analysed in.

As mentioned previously, future work can also investigate opportunities of using top2vec to improve information retrieval in several ways: (1) catch and group trending terms and slogans to improve covered content and, (2) target specific hashtags and terms with specific information.

The current study was based on an assumption of the homogeneity of the dataset when extracting tweets using 'scamdemic' and 'plandemic' words, while there is a possibility of the presence of tweets in this dataset that do not spread misinformation. In this connection, misinformation detection on short texts should be applied in advance. Recently, more sophisticated deep learning algorithms appear to tackle this problem, including Convolutional Recurrent Neural Network (Islam, 2020).

Future work could also involve extending the data from other types of sources from WHO, for example there were found collections of research documents and transcripts from informational videos. Other trustworthy sources of information could also be included in the dataset. Working with longer documents it could be approporiate to investigate methods for dividing them or match specific segments.

10. References

Angelov, Dimo. (2020). TOP2VEC: DISTRIBUTED REPRESENTATIONS OF TOPICS. https://arxiv.org/pdf/2008.09470.pdf

Baines, A., Ittefaq, M., & Abwao, M. (2021). #Scamdemic, #Plandemic, or #Scaredemic: What Parler Social Media Platform Tells Us about COVID-19 Vaccine. Vaccines, 9(5), 421.
https://doi.org/10.3390/vaccines9050421

Blei, D. M., Ng, A. Y., Jordan, M. I..(). Latent Dirichlet Allocation. https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf?ref=https://githubhelp.com

Bruns, A., Hurcombe, E., & Harrington, S. (2021). Covering Conspiracy: Approaches to Reporting the COVID/5G Conspiracy Theory. Digital Journalism, 1–22. https://doi.org/10.1080/21670811.2021.1968921

CDC Museum COVID-19 Timeline. (2022, January 5). Centers for Disease Control and Prevention. https://www.cdc.gov/museum/timeline/covid19.html
COVID-19 misleading information policy. (2021, December 15). Twitter. https://help.twitter.com/en/rules-and-policies/medical-misinformation-policy

Dimo Angelov. (2022) Documentation Top2Vec.

Douglas, K. M. (2021). COVID-19 conspiracy theories. Group Processes & Intergroup Relations, 24(2), 270–275.
https://doi.org/10.1177/1368430220982068

Egger R, Yu J. A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts. Front Sociol. 2022 May 6;7:886498. doi: 10.3389/fsoc.2022.886498.

Fuchs, C. (2021). Communicating Covid-19: Everyday Life, Digital Capitalism, and Conspiracy Theories in Pandemic Times (Societynow). Emerald Publishing Limited.

Hamid, A., Sheikh, N., Said, N., Ahmad, K., Gul, A., Hassan, L., Al-Fuqaha, A. (2020). Fake News Detection in Social Media using Graph Neural Networks and NLP Techniques: A COVID-19 Use-case. arXivLabs.
https://doi.org/10.48550/arXiv.2012.07517

Honora, A., Wang, K. Y., & Chih, W. H. (2022). How does information overload about COVID-19 vaccines influence individuals' vaccination intentions? The roles of cyberchondria, perceived risk, and vaccine skepticism. Computers in Human Behavior, 130. https://doi.org/10.1016/j.chb.2021.107176

Islam, M. R., Liu, S., Wang, X., Xu, G. (2020). Deep learning for misinformation detection on online social networks: a survey and new perspectives. Social Network Analysis and Mining, 10(1).
https://doi.org/10.1007/s13278-020-00696-x

Lanier, H. D., Diaz, M. I., Saleh, S. N., Lehmann, C. U., Medford, R. J. (2022). Analyzing COVID-19 disinformation on Twitter using the hashtags scamdemic and plandemic: Retrospective study. PLOS ONE, 17(6), e0268409.
https://doi.org/10.1371/journal.pone.0268409

Mackey, T., Purushothaman, V., Li, J., Shah, N., Nali, M., Bardier, C., Liang, B., Cai, M., Cuomo, R. (2020). Machine Learning to Detect Self-Reporting of Symptoms, Testing Access, and Recovery Associated With COVID-19 on Twitter: Retrospective Big Data Infoveillance Study. JMIR Public Health and Surveillance, 6(2), e19509.
https://doi.org/10.2196/19509

Magliani, Federico & Fontanini, Tomaso & Fornacciari, Paolo & Manicardi, Stefano & Iotti, Eleonora. (2016). A Comparison between Preprocessing Techniques for Sentiment Analysis in Twitter.

Manning, C. D., Raghavan P. and Schütze H.. (2008). Introduction to Information Retrieval (draft), Cambridge University Press. ISBN: 0521865719

Mimno, D., Wallach, H., Talley, E., Leenders, M., McCallum, A. (2011). Optimizing Semantic Coherence in Topic Models. Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, 262–272.
https://aclanthology.org/D11-1024

Nigam, A., Jaiswal, P., Sundar, S., Poddar, M., Kumar, N., Dernoncourt, F., Celi, L. A. (2021). NLP and Deep Learning Methods for Curbing the Spread of Misinformation in India. The International Journal of Intelligence, Security, and Public Affairs, 23(3), 216–227.
https://doi.org/10.1080/23800992.2021.1975429

PAHO. (2020). Understanding the infodemic and misinformation in the fight against COVID-19. https://iris.paho.org/bitstream/handle/10665.2/52052/Factsheet- infodemic_eng.pdf?sequence=16

Paul, M. J., Dredze, M. (2014). Discovering Health Topics in Social Media Using Topic Models. PLoS ONE, 9(8), e103408. https://doi.org/10.1371/journal.pone.0103408

Pertwee, E., Simas, C., & Larson, H. J. (2022). An epidemic of uncertainty: rumors, conspiracy theories and vaccine hesitancy. Nature Medicine, 28(3), 456–459. https://doi.org/10.1038/s41591-022-01728-z

Pummerer, L., Böhm, R., Lilleholt, L., Winter, K., Zettler, I., & Sassenberg, K. (2021). Conspiracy Theories and Their Societal Effects During the COVID-19 Pandemic. Social Psychological and Personality Science, 13(1), 49–59. https://doi.org/10.1177/19485506211000217

Rossi, S. (2022). The Scamdemic Conspiracy Theory and Twitter's Failure to Moderate COVID-19 Misinformation. Hawaii International Conference on System Sciences (HICSS).
Top2Vec Documentation. https://top2vec.readthedocs.io/en/latest/Top2Vec.html#benefits

Wang, Y., McKee, M., Torbica, A., Stuckler, D. (2019). Systematic Literature Review on the Spread of Health-related Misinformation on Social Media. Social Science Medicine, 240, 112552. https://doi.org/10.1016/j.socscimed.2019.112552
Wegrzyn-Wolska, Katarzyna & Bougueroua, Lamine & Yu, Haichao & Zhong, Jing. (2016). Explore the Effects of Emoticons on Twitter Sentiment Analysis. 65-77. 10.5121/csit.2016.61006.

Yang, Sidi & Zhang, haiyi. (2018). Text Mining of Twitter Data Using a Latent Dirichlet Allocation Topic Model and Sentiment Analysis.

Appendix A. Frequency of use of the most popular hashtags

Table 2: Frequency of use of the most popular hashtags

Hashtag	Frequency	Hashtag	Frequency
#scamdemic	19663	#fauciemails	472
#plandemic	15439	#psyop	445
#covid19	4142	#plandemia	428
#bihar_needs_teachers	2880	#nuremberg2	402
#covid	1664	#thegreatreset	402
#nwo	1640	#agenda2030	369
#wakeup	1164	#lockdown	352
#covidiots	1024	#event201	347
#coronavirus	903	#billgates	337
#vaccine	847	#msm	325
#greatreset	745	#vaccines	325
#freedom	715	#depopulation	318
#truth	646	#who	299
#fauci	643	#genocide	295
#covid1984	630	#scamdemic2021	285
#covidhoax	626	#newworldorder	270
#crimesagainsthumanity	599	#medicalgenocide	265
#justsayno	554	#novaccinepassportsanywhere	253
#hoax	550	#mrna	253
#agenda21	548	#mindcontrol	250
#novaccinepassport	544	#pcr	239
#canada	533	#odysee	235
#novaccinepassports	513	#toronto	230
#pandemic	509	#uk	230
#maskoff	490	#scam	230
#covidvaccine	479	#masks	228

Appendix B. LDA topics

Appendix B.1. Topics extracted by LDA topic modeling

Table 3: Topics extracted by LDA topic modeling

No	Topics	Top 20 most important words
0	Antigovernment conspiracy	vaccine, control, world, fear, government, order, change, global, human, people, system, experimental, mass, population, mind, propaganda, climate, push, agenda, passport
1	Who benefits from covid Conspiracy	money, business, government, make, making, made, health, lockdown, million, public, mark, care, paid, small, open, billion, give, pharma, politician, company
2	Punishment for “creating” covid	fauci, gate, bill, scam, crime, wuhan, virus, humanity, nuremberg, fraud, criminal, coming, evil, biggest, history, trial, research, crimesagainsthumanity, involved
3	Evidence that Covid is a conspiracy	conspiracy, fact, video, watch, doctor, anti, science, read, truth, theory, call, word, real, wrong, show, news, post, called, full, listen
4	Antivaccine conspiracy	covid, 19, vaccine, agenda, freedom, hoax, coronavirus, truth, lockdown, canada, wakeup, 1984, 21, greatreset, psyop, tyranny, pandemic, 2021, covidiots, justsayno
5	Rhetoric and language	medium, people, skin, hand, fake, light, pointing, index, tone, wake, news, make, government, sheep, woman, point, sense, good, backhand, dark
6	Antimask conspiracy	mask, people, stop, test, work, lockdown, wear, child, wearing, vaccinated, false, sick, social, fear, bullshit, school, positive, stupid, make, stay
7	Totalitarian conspiracy	pandemic, world, great, plan, planned, part, year, reset, event, real, knew, time, month, thing, game, coming, called, back, play, remember
8	Rhetoric and language	people, time, life, year, back, thing, long, good, started, family, friend, normal, work, live, feel, start, lost, week, today, living
9	Politics	trump, state, china, country, election, biden, america, american, free, democrat, fraud, party, power, economy, left, president, open, vote, liberal, communist
10	Mortality due to covid	virus, covid, death, people, vaccine, case, pandemic, year, number, died, rate, test, hospital, variant, india, million, dying, disease, deadly, fake
11	Emoticons	face, tear, floor, rolling, thinking, joyface, laughing, smiling, laughingrolling, shit, fucking, clown, symbol, fuck, mouth, yeah, medical, heart, faceclown, bullshit

Appendix B.2. Intertopic Distance Map

Figure 8: Intertopic Distance Map

Appendix C. Top2Vec

Appendix C.1. Hierarchical reduction

```
Reduced topic 1 composed of 21 initial topics
Reduced topic 2 composed of 66 initial topics
Reduced topic 3 composed of 20 initial topics
Reduced topic 4 composed of 23 initial topics
Reduced topic 5 composed of 24 initial topics
Reduced topic 6 composed of 37 initial topics
Reduced topic 7 composed of 25 initial topics
Reduced topic 8 composed of 28 initial topics
Reduced topic 9 composed of 9 initial topics
Reduced topic 10 composed of 15 initial topics
Reduced topic 11 composed of 23 initial topics
Reduced topic 12 composed of 24 initial topics
```

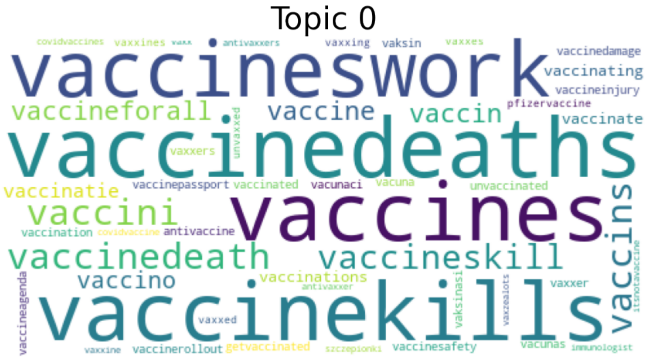
Appendix C.2. Topics extracted by top2vec topic modeling

Table 4: Topics extracted by top2vec topic modeling

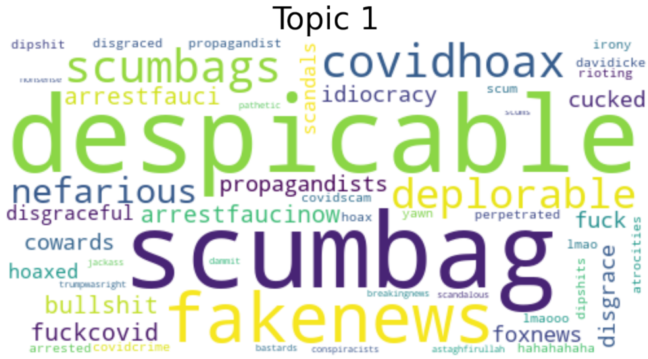
No	Topic	Top 20 most important words
0	Antivaccine conspiracy	vaccinedeaths, vaccinekills, vaccines, vaccineswork, vaccinedeath, vaccini, vaccins, vaccineskill, vaccineforall, vaccine, vaccin, vaccino, vaccinatie, vaccinations, vaccinating, vaccinate, vaccineinjury, vaccination, vaccineagenda, vaccinated
1	Rhetoric and language, media conspiracy	despicable, scumbag, fakenews, scumbags, covidhoax, deplorable, nefarious, arrestfaucinow, fuckcovid, arrestfauci, bullshit, propagandists, disgrace, idiocracy, cowards, cucked, foxnews, fuck, disgraceful, hoaxed
2	Politics, negative opinion about democrats	democrats, electionfraud, voterfraud, dems, trumpwasright, republicans, liberals, reelected, politicised, cucked, democrat, demoncrats, libtards, fearmongering, electorate, unelected, conservatives, election, dictators, trumps
3	Rhetoric and language	bullshit, scumbag, yawn, bastards, fuck, hahahahaha, lmao, fuckfauci, scumbags, dipshit, dumbass, hahahahahaha, goddamn, furiously, lmaooo, nefarious, dipshits, idiots, idiocy, wow
4	Vaccine and epidemic conspiracy	virushoax, epidemic, vaccinedeaths, epidemics, vaccinedeath, coronavirushoax, influenza, pandemic, flu, plagues, vaccinekills, panicdemic, plague, eatthevirus, hoaxdemic, coronavirus, outbreak, falsapandemia, coronaviruses, scaredemic
5	Rhetoric and language	kidding, dammit, chums, furiously, whoops, deny, fuckthenwo, damned, waaaay, seriously, bother, blatantly, whoa, succumbed, skyrocketed, hastily, damning, unagenda, staggering, bothered
6	Antimask	maskface, facemasks, facemask, fuckmasks, mask, fuckyourmask, maskless, masks, unmasking, masksoff, unmasked, masked, maskmandates, maskprohibited, wearamask, maskoff, nomask, nomasks, maskholes, maskhole
7	Information channels	truthers, fakescience, hoax, sciencefraud, truther, fakenews, faux, hoaxed, fake, bullshit, untrue, fraudfauci, fraudulently, verdade, frauds, faked, conspiracists, truthwins, verdad, covidhoax
8	Globalism	globalists, globalist, globaldemocracy, globalism, worldnews, totalitarian, mundo, wereld, world, globalwarming, globalminions, globalnews, globalreich, globally, globale, dunia, wereldwijd, globalreset, mundial, worldwide
9	Socioeconomic differences	bankrupt, bagmoney, poors, greed, greedy, billionaires, profited, profits, scumbags, poor, capitalists, bankruptcy, richest, rich, frauds, recession, wealthiest, wealthy, economically, wealthier
10	Antimask and antivaccine, Rhetoric and language	fuckyourmask, unvaxxed, takeoffyourmask, malicious, nefarious, deplorable, premeditated, frighten, nomask, idiocy, despicable, atrocities, panicked, arrestfaucinow, epidemic, vaccinedeaths, fuckmasks, denying, cowards, nonsense
11	Rhetoric and language	already, previously, skyrocketed, meanwhile, antes, untill, eventually, earlier, afterall, lastly, til, upcoming, latter, ultimately, lasted, ended, indefinitely, overdue, skyrocketing, lately

Appendix C.3. Top2vec wordclouds

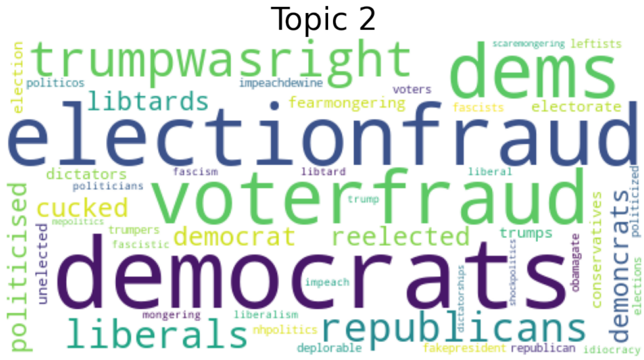
Figure 9: Top2Vec word-clouds



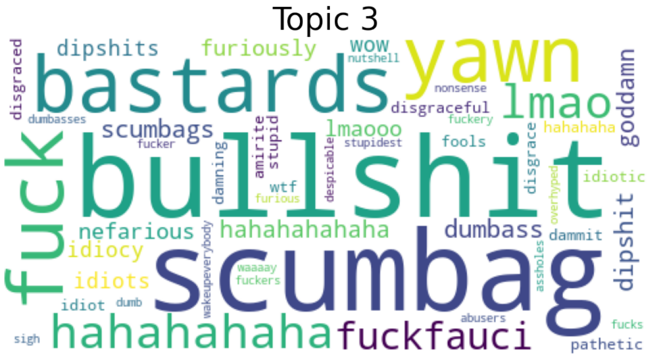
((a)) Anti-vaccine cosnpiracy



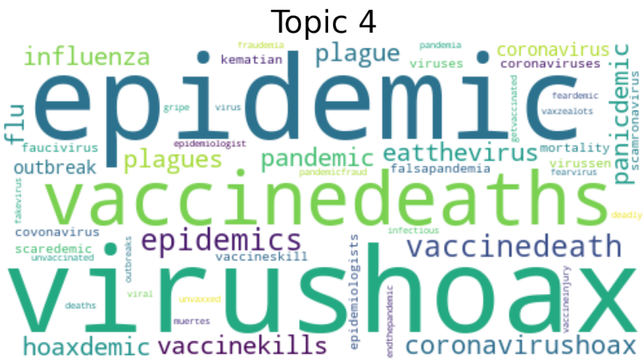
((b)) Expressive rhetoric and language, conspiracy



((c)) Politics



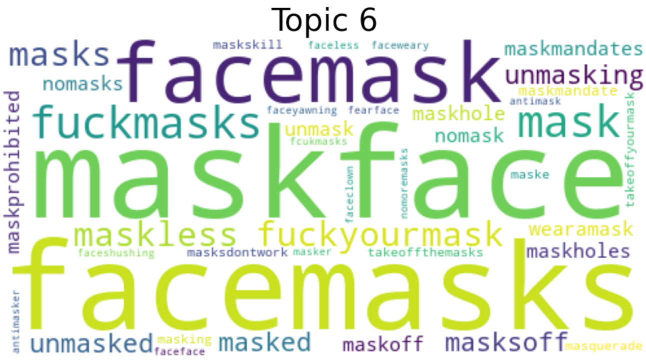
((d)) Expressive and humiliating rhetoric and language, slander



((e)) Anti-pandemic and anti-vaccine conspiracy and

((f)) Expressive rhetoric and language

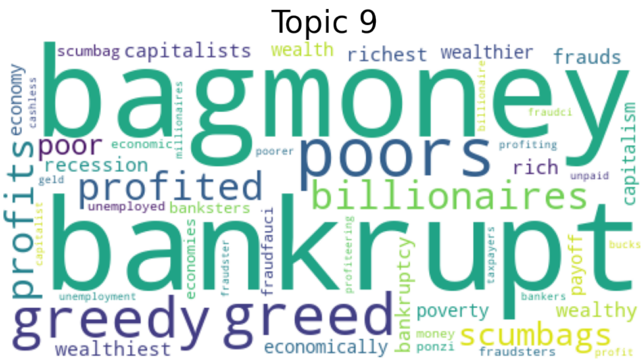
Figure 10: Top2Vec word-clouds



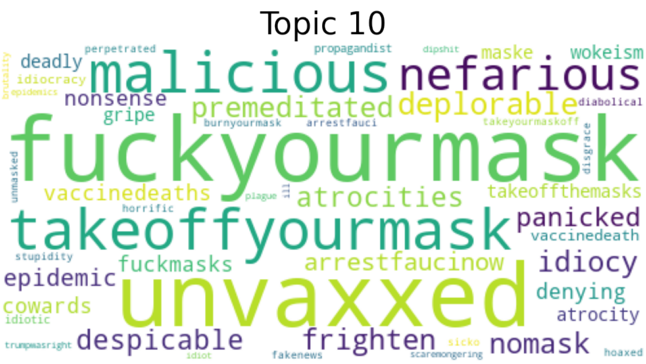
((a)) Mask refusal

((b)) Distrust in mainstream information channels and science

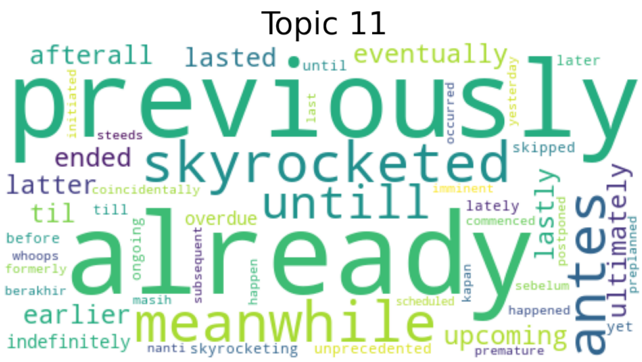
((c)) Globalisation perspective with slight anti-government references



((d)) Socioeconomic topic with seemingly negative characterizations.



((e)) Anti-mask and anti-vaccine opinions, slander



((f)) Seemingly neutral rhetoric and language

Appendix D. Information Retrieval

Appendix D.1. Retrieved documents for Top2vec topics

	Topic	WHO_docID	Score	Document
0	vaccinedeaths vaccinekills vaccines vaccineswo...	395	9.082741	all vaccines with who emergency use listing ar...
1	despicable scumbag fakenews scumbags covidhoax...	None	0.000000	
2	democrats electionfraud voterfraud dems trumpw...	240	9.407431	as of march 2020, there have been reports of t...
3	bullshit scumbag yawn bastards fuck hahahahaha...	None	0.000000	
4	virushoax epidemic vaccinedeaths epidemics vac...	322	10.729904	1. treatments for covid-19 and influenza are d...
5	kidding dammit chums furiously whoops deny fuc...	64	4.503864	decisions about mask use in children should be...
6	maskface facemasks facemask fuckmasks mask fuc...	257	4.775566	how to put on and wear a fabric mask:before to...
7	truthers fakescience hoax sciencefraud truther...	None	0.000000	
8	globalists globalist globaldemocracy globalism...	514	6.571720	who member states have developed multiple glob...
9	bankrupt bagmoney poors greed greedy billionai...	234	4.223078	yes, any situation in which people are in clos...
10	fuckyourmask unvaxxed takeoffyourmask maliciou...	64	4.503864	decisions about mask use in children should be...
11	already previously skyrocketed meanwhile antes...	518	4.749197	as the process of the intergovernmental negot...

Figure 11: Retrieved documents for Top2vec topics

Appendix D.2. Retrieved documents for LDA topics

	Topic	WHO_docID	Score	Document
0	vaccine control world fear government order ch...	499	17.882354	who'score normative function is to compile and...
1	money business government make making made hea...	143	13.765360	for most decisions, multiple ethical values an...
2	fauci gate bill scam crime wuhan virus humanit...	477	11.116912	sars-cov-2 spreads primarily through human-to-...
3	conspiracy fact video watch doctor anti scienc...	36	13.110766	a near-constant stream of news, sometimes cont...
4	covid 19 vaccine agenda freedom hoax coronavir...	211	8.362336	large scale physical distancing measures and m...
5	medium people skin hand fake light pointing in...	494	11.205480	itis a regular process that who consults with ...
6	mask people stop test work lockdown wear child...	147	19.162754	disposable glovesgloves may be used by food wo...
7	pandemic world great plan planned part year re...	336	9.010466	try and reduce long periods of time spent sitt...
8	people time life year back thing long good sta...	334	17.566940	who has detailed recommendations on the amount...
9	trump state china country election biden ameri...	470	13.431620	violence against women is highly prevalent, an...
10	virus covid death people vaccine case pandemic...	483	23.927673	while aggregate covid-19 case and death number...
11	face tear floor rolling thinking joyface laugh...	258	12.344752	how to put on and take off a medical mask:befo...

Figure 12: Retrieved documents for LDA topics

Appendix D.3. Retrieved documents for tweets subset

	Tweet	WHO_docID	Score	Document
0	breaking brexit was waste time gives franc...	147	10.860115	disposable glovesgloves may be used by food wo...
1	another not wearing gloves	147	8.453755	disposable glovesgloves may be used by food wo...
2	vernon coleman final irrefutable proof that ...	209	6.518293	attempts to reach 'herd immunity' through expo...
3	manufacturedcrisis psyop 4ir commoditizati...	84	8.454457	there is no evidence of a direct connection be...
4	planet ccp can check out anytime want comr...	470	7.082640	violence against women is highly prevalent, an...
5	notapsychicbut fauciforprison for hiding the...	311	6.299819	in countries or areas where there is intense c...
6	thankful for mochafest least some people will...	197	7.093324	workplaces for jobs at medium risk require dai...
7	also agree there much behind the scenes th...	20	6.464889	many of the initial clinical studies with covi...
8	has absolutely nothing with health and wel...	145	6.095113	health care workers (caring for patients) and ...
9	should thank him thought was fake virus ...	474	7.295954	when a virus replicates or makes copies of its...
10	yes like say lot people don have anyth...	406	13.594744	yes. the maximum level of protection from covi...
11	yep tells you everything you need know about...	309	5.781562	in a situation like this it is normal to feel ...
12	never once wore mask any stage this wor...	339	9.753227	there is no evidence that the virus that cause...
13	seen the discrepancies and the doubt	370	4.753399	if you become ill during your travel, inform y...
14	not designed affect the youth especially asi...	46	7.596576	the anonymity of working from home is really t...
15	nwo depopulation dictators desperate push t...	87	10.320639	the covid-19 pandemic is a public health emerg...
16	parts amp	60	3.891415	these drugs (casirivimab and imdevimab) are an...
17	the facts the crimes against humanity laws...	46	6.138744	the anonymity of working from home is really t...
18	yep pathetic lebron says nothing china oppre...	36	5.386446	a near-constant stream of news, sometimes cont...
19	using the pandemic scam the public purse o...	384	8.668763	who is one of the leaders of a global alliance...
20	variants are this summers masks something ke...	325	8.666914	the most effective way to protect yourself fro...
21	are this situation because our govt doesn ...	298	12.968470	losing someone close to you is always hard, wh...
22	zany face	73	4.494621	in the context of covid-19, some children may ...
23	because the experimental covid injection stil...	387	8.187444	in a regular vaccine study, one group of volun...
24	the president need address loadshedding issue...	426	5.538147	in rare situations where a serious adverse rea...
25	funnily the richest doubled their worth during	492	4.570779	every model is an approximation of reality. mo...
26	true great has left alanwatt please list...	36	10.001501	a near-constant stream of news, sometimes cont...
27	when real world vaccine trials wrong people ...	395	6.105629	all vaccines with who emergency use listing ar...
28	had falling out with parents prior rona ov...	42	8.169626	this is a difficult time. many people – includ...
29	since the beginning the wore mask less t...	455	6.575827	in the context of the covid-19 pandemic, there...

Figure 13: Retrieved documents of random tweet subset