Using Network Analysis and Natural Language Processing Methods to Study Dark Side of Social media with Applications in Fake news and Anti-vaccine Messages

by

Alireza Farnoush

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
August 6, 2022

Keywords: Fake News, COVID-19, Recommendation Systems, Natural Language Processing, Network Analysis, Anti-vaccine

Approved by

Gregory Purdy, Chair, Assistant Professor of Industrial & Systems Engineering, Auburn University
Ashish Gupta, Co-chair, Professor of Raymond J. Harbert College of Business, Auburn University
Konstantinos Mykoniatis, Assistant Professor of Industrial & Systems Engineering, Auburn University
Han Li, Associate Professor of Anderson School of Management, University of New Mexico
Amit Morey, Associate Professor of Poultry Science, Auburn University

Abstract

Social media has revolutionized the way people consume information and communicate with each other. Despite the several benefits and opportunities of the advent of social media, such as crisis management and situational awareness, social media's dark side has affected people's lives and society. One of the primary concerns regarding social media is shaping the public's perception and opinion toward specific topics. Two examples of social media's dark side are the dissemination of fake news and anti-vaccine messages. The severe impacts of fake news and anti-vaccine messages can be enormous in situations such as the COVID-19 pandemic, where people's behavior plays a significant role in managing this pandemic.

In this dissertation, we study fake news and anti-vaccine message. The first goal of this study is to design a recommendation system that detects and filters out fake and satirical news and recommends only real news. To develop a real news recommendation system, first, we built a fake and satirical news detection model by training a random forest on the distribution of topics in each article. The distribution of topics has been extracted using topics modeling techniques: latent Dirichlet allocation and latent semantic analysis. The model achieves an accuracy of 85% in recognizing fake and satirical news from real. We also used a lexicon-based sentiment analysis model to extract the sentiment of articles. Second, a K-nearest neighbor finds the K similar real news based on the distribution of topics and sentiment similarity of users' latest read news.

The second goal of this dissertation is to narrow the fake news studies, where we focus on fake news in the context of COVID-19. We aim to check and study whether deception theories can help reveal the strategies used by fake news writers in COVID-19 to deceive the audience. We used natural language processing techniques and partial least squares structural equation modeling

analysis to measure and test the strategies. Further to evaluate the results, we built a detection model by applying XGBoost to the discovered deceiving strategies. The results suggest interesting findings. For example, We found that fake news writers in the context of COVID-19 use significantly more uncertain language, more negative affect, less diversity, more expressive words, and more cognitive process in their writing.

The final goal of this dissertation is to study anti-vaccine posts on Facebook. We first seek to check if there are heterogenous topical groups of posts related to the COVID-19 vaccine on Facebook. Then we intend to study and contrast the anti-vaccine group with other discovered groups in terms of emotion and network characteristics. We implement a semantic network based on semantic similarities between the posts. To find semantic similarity, we integrate a BERT model with the cosine-similarity method. We found five giant topical groups and named them based on the major topics in each group. The results of emotion analysis show higher emotional posts and more negative emotions in anti-vaccine groups. Also, the network characteristics of groups indicate political and anti-vaccine are more topical homophily and target a specific audience.

Acknowledgments

I am deeply thankful to my advisor, Dr. Greg Purdy, for his support, patience, and guidance during my Ph.D. I am extremely indebted to my co-advisor, Dr. Ashish Gupta, for his endless assistance, advice, and training in making me a researcher. Without his resources and guidance, I would have not reached this point. I wish to acknowledge Dr. Han Li for mentoring and training me. Thanks also to Dr. Konstantinos Mykoniatis and Dr. Amit Morey for spending time reviewing my dissertation and providing me with their feedback and encouragement.

In addition, I would like to express my deepest gratitude and love to my parents and sisters for their love, support, and inspiration. Without having their support throughout my life, I would not have been able to achieve this goal. I'm very grateful to my wife for being patient during this Ph.D. process. Finally, I would like to express my sincere love for my son. He has brought a lot of energy and motivation to me by coming into my life.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

NLP	Natural Language Processing
LDA	Latent Dirichlet Allocation
LSA	Latent Semantic Analysis
BOW	Bag of Words
RF	Random Forest
BERT	Bidirectional Encoder Representations from Transformers
PLS-SEM	Partial Least Squares Structural Equation Modeling

1 Introduction and Motivation

1.1 Introduction

Today, the traditional form of news is no longer the primary source of information. People can access and share news with others, easier and faster than ever before. A study by (Jin et al. 2014) showed that young adults use social media frequently in their daily lives, which shows how popular it is. These characteristics of social media provide valuable opportunities, such as crisis management and situational awareness. For example, during Storm Cindy in 2017, social media was used by weather and governmental agencies and the public to diffuse emergency information (Kim et al. 2018).

On the other hand, social media are misused to shape public opinion toward specific topics. For example, using social media in propagating wrong information to change the public opinion about the 2016 election and climate change have been studied by Bessi and Ferrara (2016) and Samantray and Pin (2019), respectively. Another example of the most concerning situation in which social media was misused to change public opinion is the COVID-19 pandemic. Fake news writers and anti-vaccine groups are two groups that misuse social media to disseminate their views and affect others during COVID-19.

In this dissertation, we aim to use analytical approaches to understand and fight against fake news and anti-vaccine messages. To this end, we first realized that social media's recommendation system could play a significant role in the fight against fake news. The previous research either developed a fake news detection model or recommendation system, and no study covers both under one topic. Combining both a fake news detection and recommendation system can significantly contribute to the fight against fake news. Second, while we focused on fake news related to COVID-19, we found a gap where few previous studies developed a theory-based model to investigate fake news related to COVID-19. Therefore, we extend different deception theories to decipher the strategies that fake news writers of COVID-19 might use in their writings. Finally, we found the third gap in the previous research that examine the messages that anti-vaccine groups share on Facebook. Few previous studies created the semantic network to find the anti-vaccine groups on Facebook. In addition, no studies investigated their network characteristics.

The analytical methods used in this study heavily rely on natural language processing (NLP), network analysis and machine learning methods. Utilizing NLP and machine learning methods to extract textual features is very common in fake news detection studies (Saquete et al. 2020). More specifically, in this dissertation, we utilized topic modeling as an NLP technique, random forest, XGBoost and partial least squares structural equation modeling (PLS-SEM) as machine learning techniques, and semantic and co-occurrence network as network analysis methods.

1.2 Contribution

The contribution of this study is three-fold. In contribution 1 (chapter 2), we defined fake news and discussed social media's role in disseminating fake news. Then we suggest a recommendation system model that fights against spreading fake news through social media. In the second contribution, chapter 3, we contextualized the problem of fake news to a specific catastrophe event such as the COVID-19 pandemic, where fake news with specific characteristics and thematic focus associated with COVID-19 is created. In this part, we provided the theoretical foundation to develop a better understanding of the science of fake news. Inspired by deception theories, we raised the strategies used by fake news producers and investigated the significance of those strategies. In the third contribution, we investigated and classified the Facebook posts regarding the COVID-19 vaccine shared on public pages. Specifically, we compared and studied the anti-vaccine topical group with other topical groups on Facebook.

A real news recommendation system based on topic and sentiment similarities: Social media platforms unwantedly help fake news be shared million times and spread much quicker than before. Moreover, social media's recommendation systems are blamed for creating an echo chamber and filter bubble, which contribute to the dissemination of fake news and cause a false sense of validation in the readers. To fight the dissemination of fake news on social media, we proposed a two-stage recommendation system in which, in the first stage, the model can detect types of news, namely fake, satire, and real. In the second stage, the model recommends similar real news to the last news article that a user has read. Our suggested recommendation system investigates the similarities in two layers of filters. In the first filter, similarities of latent topics (topic probability distribution) in the content of articles are studied. In the second filter, the sentiment similarities of content are analyzed. In this study, we combined topic modeling and Random Forest to detect the type of news and applied the KNN model to find similarities among news based on the topics and sentiment.

Analyzing the language used in fake news related to COVID-19: In the eras of "fake news" and "social media," fake news producers take advantage of any crisis to produce and disseminate fake news. The flood of fake news in a situation such as the COVID-19 pandemic, which affects many countries and takes many lives, can increase societies' fear and stress and mislead people in fighting this virus. In this study, inspired by different deception theories, we raised hypotheses to analyze the language used in fake news and compare it to the real news associated with COVID-19. To test the hypothesis, first, we discussed the linguist cues that can help measure the hypotheses and leveraged NLP techniques to measure the linguistic cues. Second, we implemented the PLS-SEM method to examine the differences between fake news and real news related to COVID-19. In addition, we used multivariate analysis of variance (MANOVA) to support our findings.

Analyzing the characteristics of Facebook posts about COVID-19 vaccine: While COVID-19 vaccines have proven effective, many people still refuse or delay accepting the vaccine. Reading information through social media can affect people's decisions about the vaccine. In this contribution, we analyzed the Facebook posts semantically through Bidirectional Encoder Representations from Transformers (BERT). Then we created a semantic network and implemented Louvain algorithm methods to classify Facebook posts into different topical groups. Subsequently, we compared and contrasted the topical groups in terms of network characteristics and emotion. Unlike the previous studies that mostly used lexicon-based methods to analyze the emotions, we leveraged a Bidirectional Encoder Representations from Transformers (BERT) model.

The remainder of this dissertation is structured as follows. In chapter 2, we introduce our research about fake news. In this work, we developed a real news recommendation system. Our study about Fake news related to COVID-19 is presented in chapter 3. Chapter 4 presents our analysis of the characteristics of anti-vaccine posts on Facebook. Finally, Chapter 5 includes our conclusions and future work.

2 A real news recommendation system based on topic and sentiment similarities

2.1 Introduction

The growth of the internet and the advent of social media has dramatically changed the news consumption model. Nowadays, social media overtakes traditional print newspapers and has become the primary source of news for young Americans (Shearer 2018). Social media broadly refers to forums, social networks, blogs, microblogs, social news, wiki, etc.

Although the advent of digital news provides users with easy access to a large amount of information from different sources (information overload), it is hard for users to browse their news of interest (Feng et al. 2020). Recommendation systems can address the information overload and assist users in finding their information of interest based on factors such as their consumption history (Ribeiro et al. 2014). A good recommendation system provides the most relevant information to the user at the right time and place (Shokeen and Rana 2020). Historically, many organizations in different domains used recommendation systems to recommend an item to the customers (Natarajan and Moh 2016, Feng et al. 2020).

One of the controversial recommendation systems was news recommendation systems used by social media, where social media use the user's profile and activities to recommend news (Ashraf et al. 2018). For example, a news recommendation system on Twitter might extract features from users' tweets, and Facebook can benefit from users' comments and pages like to analyze their interests and recommend news (Ashraf et al. 2018). It is prudent to believe that social media news recommendation systems can play an essential role in disseminating the news, in the era of digital news when social media is a source of information.

Moreover, the nature of social media, where people can easily share any news, makes it a good target for spreading misinformation. For example, Twitter and Facebook have been blamed for spreading misinformation and fake news, especially during the 2016 election (Lex et al. 2018). Malicious bots and social media users can generate and spread fake news. Surprisingly, even social media's news feed assists in facilitating the spread of fake news by creating echo chambers and filter bubbles (Mohseni and Ragan 2018). The echo chamber is a situation in which a person is exposed to specific Information (Jamieson and Cappella 2008); therefore, it increases the chance of accepting fake news and false view by increasing the false sense of validation (Mohseni and Ragan 2018). The filter bubble is a similar theory, and it is a digital echo chamber that was first defined by (Pariser 2011) as a state of intellectual isolation. A filter bubble happens when users see selective and tailored content consistent with their pre-existing view (Geschke et al. 2019). Filter bubbles confine the news diversity by filtering out the information that is not classified as a user's interest (Haim et al. 2018). Therefore, owing to a news feed system, when a person is exposed to an extreme left or right view or a fake news story in social media, it is unlikely to see other news sources to challenge the accuracy of the story by seeing other political views or other versions of the story (Sindermann et al. 2020).

Nowadays, millions of users visit social media daily. The latest figures show that Facebook and Twitter have more than 1.82 billion and 152 billion daily active users, respectively (Sehl 2019, Aslam 2020). Consequently, fake news stories can be quickly published and shared millions of times by users on social media. Some examples of posting fake news on social media include sharing fake news articles from denverguardian.com website about the nominees in the 2016 US presidential election on Facebook (Allcott and Gentzkow 2017) or the fake news related to the infamous Pizzagate story published on Facebook, Twitter, and Reddit (Lopez 2016). The widespread of such fake news stories raised concerns about the adverse effect of fake news on society and individuals (Zhang and Ghorbani 2020). The urgent need to detect fake news and mitigate the spread has drawn attention from academia and industry.

The first step to combating fake news in social media and avoiding creating a filter bubble and echo chamber is to understand the life cycle of fake news in social media and know where in the life cycle we want to fight. The life cycle of fake news in social media has three phases: creation, publication, and propagation (Zhou and Zafarani 2018a, Mohseni and Ragan 2018). Fake news can be blocked in each of these phases. In this study, we detect fake news in the creation phase and mitigate the spread in publication phase.

Despite the influential role of the news recommendation system in spreading fake news, most researchers investigate fake news and news recommendation system independently. Few studies examine the fake news detection and news recommendation system under one topic. In this study, we designed a news recommendation system that can detect and filter out fake and satire news and recommend N real news based on the user's latest news reading. Our model combats fake news in two stages (out of three stages) of a fake news life cycle. First, our methods can detect fake news in the creation stage. Then, the technique fights in the second stage (publication) by mitigating the spread of fake news. Moreover, the proposed model reduces the likelihood of creating the filter bubble and echo chamber.

The remainder of this contribution is structured as follows. In section 2.2 (related work), we introduce some state of the art related to this topic. In sections 2.3 and 2.4, we cover the methods used study and the experiment setting. In sections 2.5 and 2.6, the results and discussion are presented, respectively. Finally, we present the conclusion in section 2.6.

2.2 Related work

2.2.1 Fake news

Although fake news is not a new concept, its usage and meaning have changed in the last decade (Sharma et al. 2019). The current form of fake news was raised during the 2016 US presidential election when fake news about the candidates spread through social media (Allcott and Gentzkow 2017, Barthel et al. 2016). Fake news was defined in different ways. Allcott and Gentzkow (2017) defined fake news as “news articles that are intentionally and verifiably false and could mislead readers,” and Golbeck et al. (2018) defined it as “information, presented as a news story that is factually incorrect and designed to deceive the consumer into believing it is truth.” Two important attributes are common in most fake news definitions: intention and falsehood.

Rubin et al. (2015) categorized fake news into three categories: serious fabrication, hoax, and satire. Another version of false news is satire. Unlike fake news, satire is not designed to mislead readers, and it is assumed that both authors and readers know the humorous nature of satire (Tandoc Jr et al. 2018). However, satire might fool the reader if they do not get the latent humor behind the satire. There is another view about satire that emphasizes that the humor nature of satire is not designed to only amuse the reader but also is used to criticize, but the humor inside the satire relieves the hardness of critique (Tandoc Jr et al. 2018). Historically, satire played a role in making doubt in the reader’s mind, shaping their opinion and stimulating the reader’s ability to judge society (Burkhardt 2017).

2.2.2 Fake news detection

Fake news producers design the news so that the readers believe it is real news (Tandoc Jr et al. 2018). Human judgment in recognizing deception is less than 63 percent (Rubin and Conroy 2012). On the other hand, artificial intelligence and machine learning showed promising results in fake news detection. Recently, many different fake news detection systems have been developed to detect fake news. All fake news detection models can be broadly grouped into two classes: content-based and propagation-based methods (Zhou et al. 2020a, Zhou and Zafarani 2019, Conroy et al. 2015).

Content-based method: the methods in this category are based on the features extracted from news content. The extracted features usually are textual and visual (Shu et al. 2019). Textual features are the focus of this study. The approaches used textual features can be classified into three categories: knowledge-based, style-based, and latent features (Zhou et al. 2020a). knowledge is constructed based on a set of (subject, predicate, and object) (SPO) extracted from the text. In this system, the truthfulness of news is obtained by comparing the knowledge extracted from to-be-verified news with ground truth (Shi and Weninger 2016, Ciampaglia et al. 2015). For example, recently, Zhang et al. (2019) proposed a fake news detection system (FEND) in which they detect fake topics and fake events. They utilized the SPO set to identify topics and events from the sentences. The style was defined as a set of quantifiable textual features (Zhou and Zafarani 2018a), including a wide range of features. For example, Reis et al. (2019) evaluated 141 textual features and classified them into five categories: syntax, lexical, psycholinguistic, semantic, and subjectivity features (check the subjectivity score of each sentence) (Reis et al. 2019). Syntax features are sentence-level features such as a bag of words, part-of-speech, n-gram, etc. (Reis et al. 2019, Zhou et al. 2020a, Feng et al. 2012, Shu et al. 2017). Lexical features are the frequency statistic of features such as unique words, verbs, first-person, etc. (Zhou et al. 2020a, Reis et al. 2019). Psycholinguistic features include the method used in Linguistic Inquiry and Word Count (LIWC), a dictionary-based software. This software can find the proportion of words in different psycholinguistic categories, such as the proportion of words in sadness, happiness, etc. Some studies categorized psycholinguistic features as a part of semantic or lexicon features. In a semantic feature, semantic similarities of news are investigated (Bharadwaj and Shao 2019, Hardalov et al. 2016, Jadhav and Thepade 2019). Latent category involves the features generated by tensor factorization (Hosseinimotlagh and Papalexakis 2018) and deep learning (Karimi et al. 2018a, Wang et al. 2018, Zhou et al. 2020b). However, the black-box nature of these methods makes them hard to interpret, though good results in detecting fake news (Zhou et al. 2020a, Shu et al. 2019).

Propagation-based methods: This method detects fake news by investigating the propagation pattern of fake news on social media (Vosoughi et al. 2018, Antoun et al. 2020) and is based on the features related to the social context information. For instance, Zhou and Zafarani (2019) found that fake news spreads faster than real news. In addition, the fake news spreaders make more engagement with the news and have a denser network. In another study, Vosoughi et al. (2018) showed that fake news spread faster, more profound, farther, and more largely than real news.

Satire detection is also part of controlling the spread of false stories as often people get fooled by satire (De Sarkar et al. 2018), and many studies have been conducted to detect satire news. Rubin et al. (2016) propose a model that was able to distinguish satire from real with 90% accuracy. Rashkin et al. (2017) discussed that satire is closer to fake news than real news regarding information quality. Horne and Adali (2017) analyze three categories of news: real, fake, and satire. They concluded that satire and fake news are more similar than fake and real in terms of complexity and style. The textual features discussed in the fake news detection model can also be utilized in the satire detection model. For example, De Sarkar et al. (2018) and Yang et al. (2017) used linguistic features in their deep learning satire detection model to detect satire. De Sarkar et al. (2018) showed that semantic features are more effective than syntax features in detecting satire. In another study, (del Pilar Salas-Zárate et al. 2017) used psycholinguistic features in their satire detection model.

The method in this study to detect fake or satirical news is style-based. The prior studies in this category mostly used different types of textual features. Despite obtaining high accuracy from the previous studies, there is a doubt about the future of this method as fake news producers are making their writing style similar to the real news. On the other hand, fake news writers focus on different topics than real news. For example, Gupta et al. (2022) discussed that the themes used in the fake news articles in the context of COVID-19 are different than real news. In this study, we based our model on the assumption that fake, satire, and real always focus on various themes. Therefore our model distinguishes between fake, satire and real based on the themes used in the news, not the feature that might become ineffective in the future or a different context.

2.2.3 Recommendation System

There are three main methods in a recommendation system: content-based filtering, collaborative filtering, and hybrid. The content-based (CB) recommendation system, as the name implies, is based on the content of an item indicating the user's preference. This system constructs a profile for a user based on the items a user has already shown interest in. Then, the user's interest is compared to the new item's content to find appropriate items to recommend to the user (Pazzani and Billsus 2007). Collaborative filtering (CF) recommends an item based on the opinion of other users (Schafer et al. 2007). Compared to CB, CF ignores items' content and recommends an item to a user based on the similarity between users. Users are similar if they are interested in the same items. The hybrid system is a combination of CB and CF. User interest can be measured implicitly like hitting or explicitly like rating score, like, and dislike reaction.

One possible challenge in the recommendation system is the lack of enough information to find customers' interests, which is called a cold start problem (Schein et al. 2002). The cold start problem happens in three situations: when a new user enters the system, when a new item is offered, or when a new community forms, which is a combination of a new user and a new item (Schafer et al. 2007).

Recommendation systems have been applied to different domains, such as recommending movies (Phorasim and Yu 2017, Lin et al. 2016, Kuzelewska 2014), books (Jomsri 2014, Mooney and Roy 2000), music (Chen and Chen 2001, Hauver and French 2001), online shopping (Linden et al. 2003) and news (Liu et al. 2010, Lu et al. 2015). News is the focus of this paper.

The news recommendation system is different from other domains in many aspects. First, in the news recommendation system, users' interest can hardly be measured since it is not normal to ask a user to rate a specific piece of news. Moreover, not all social media platforms have 'like' and 'dislike' options. Therefore, in this domain, it is assumed that reading a news article means the user is interested in that news (Cleger-Tamayo et al. 2012). Second, a news article's popularity can change in a short period relative to breaking news, so the recommender system should continuously update the information and model. Third, user interest can vary rapidly based on situational factors, such as time, location, and tools (Karimi et al. 2018b, Campos et al. 2014).

In this study, we consider no history of the users and formulate this study as a cold start problem. We want to make the recommendation to a new user who lacks history on social media. Many studies addressed the cold start problem. For example, (Tavakolifard et al. 2013) proposed a model to recommend a piece of news to a user who lacks a profile based on location similarity between that user and the news, latest preferred articles, and popularity of the news articles. Similarly, Fortuna et al. (2010) addressed the problem of a new user by leveraging content information such as the users' local time, location, and latest read news. Lee and Park (2007) used the demographic data of a new user, like age and sex, to recommend news based on the interest of that demographic profile.

2.3 Research Methods

Our analytics model occurs in two stages. In the first stage, a detection model distinguishes the type of news articles, and in the second stage, we form a news recommendation system.

In this study, we have developed a Topic Modelling_Random Forest Model. We assumed all news articles shared on social media were collected in a news repository. In the first stage, we applied two topic modeling methods, namely Latent Dirichlet Allocation (LDA) and Latent Semantic Analysis (LSA), to the content of all collected news articles in the news repository. In the next step, we convert topic probability distribution, which is the output of our topic modeling, to feature vectors. Then a supervised machine learning method (random forest) is applied to feature vectors to classify the news articles. This method is in line with prior research such as (Savov et al. 2020) where topic modeling is used to predict paper publication years. In another study, Liu et al. (2016) integrated LDA and support vector machines to classify web services. They proved that this method obtains better accuracy than the term-based method, where each term is considered a feature. They also discussed some of the advantages of this method over the term-based method, such as low sparseness. Finally, at the end of the first stage, we filter out fake and satirical news and collect real news in the legitimate news repository. Figure 2.1 shows the framework of our method.

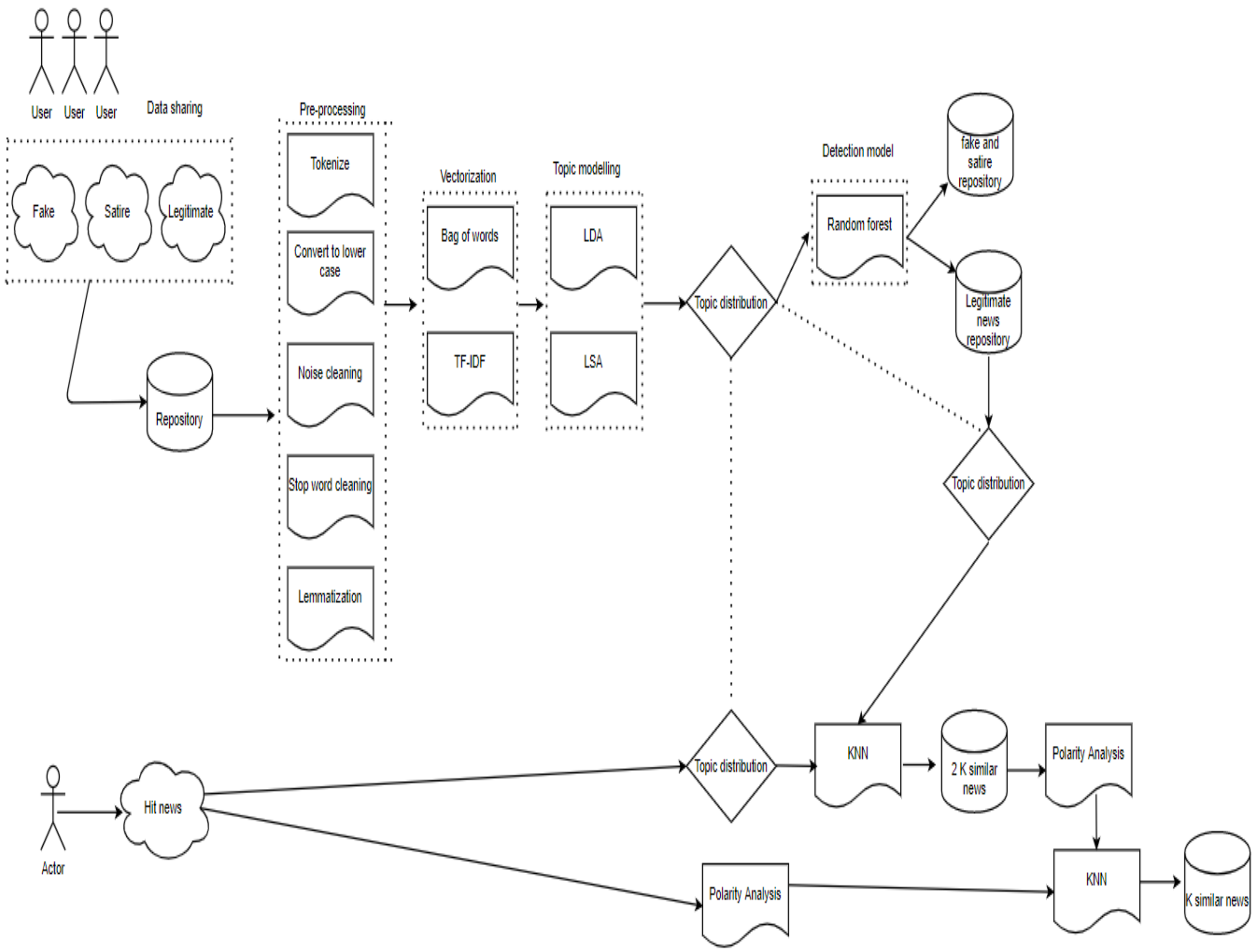

Figure 2.1: Model Overview

In the second stage, we applied a content-based filtering recommendation method to the user's last read article. We aim to recommend N real news to a user in a cold start problem where we have either a new user or a new article. As one of the possible solutions to a cold start problem, we used the latest read news by users. We based our recommendation system on the similarity of topics and sentiment. However, the topic similarity is the priority of our recommendation system. Therefore, we analyzed the similarity in two layers where in the first layer, news articles are filtered based on their topic distribution similarities to the latest read news article. In the second layer, the articles are filtered based on their sentiment similarities. We applied a k-nearest neighborhood

(KNN) method on each layer to find similar real news. In other words, in the first layer of filtering, the KNN model ingests topic distribution vectors and filters 2 N news articles from the legitimate news article repository. In the second layer, another KNN model filtered N news articles out of 2 N articles based on sentiment scores. Using KNN in a recommendation system is consistent with prior studies such as Amatriain et al. (2011). In addition, using sentiment analysis in a recommendation system is in line with previous studies, such as work by (Parizi and Kazemifard 2015, Strapparava and Mihalcea 2008). In the following, we explain all steps behind the mentioned methods.

2.3.1 Pre-Processing

We employed Spacy (Honnibal 2017) and NLTK libraries package in Python to do textual pre-processing. Text pre-processing begins with Tokenization, meaning that the whole article is turned into a collection of words where each word represents a token. Then, we transferred all words into their lowercase format and removed stop words (such as “the”, “and”, “in”). We also removed any words that reveal the source of the article. For example, we removed “NBC News” from the articles since we wanted to make sure our model was not biased. Next, we applied a lemmatization algorithm on tokens to remove the words' inflectional parts.

2.3.2 Vectorization

The next step is vectorization, which converts each document into a numerical vector based on the tokens. We applied two types of vectorizations described below. However, before vectorizing the data, we split the dataset into two sets of test and train, where we withheld 20% of the data to test the performance of our model and 80% for the training purpose.

Term Frequency-Inverse Document Frequency (TF-IDF): Term frequency (TF) is the number of occurrences of a term in a document, while inverse document frequency (IDF) is the inverse of the number of documents that include the terms. This method gives weight to the words based on their importance in a document. Therefore, a high TF-IDF weight in a document indicates the high frequency of the term in that document and the low number of documents that include the term (Ramos 2003). IDF imposes a penalty and lowers the weight of terms that appears in many documents (Schütze et al. 2008). The formula associated with IDF is shown below. In this formula, N is the total number of documents in the corpus and df_t is the number of documents containing term (t).

$$\mathrm{IDF}\ (\mathrm{t}, \mathrm{D}) = \log\ (\frac{N}{|df_t|})$$

Bag of Words (BOW): is a simple representation of the occurrences of the terms that each document contains. In contrast to TF-IDF, BOW is not sensitive to the number of documents that contain a word. We set both TF-IDF and BOW to ignore the words that appear in more than 95% of articles and less than 5% of documents, as these words have low semantic values.

2.3.3 Topic Modeling

Topic modeling helps discovers the latent topics in the articles. Each topic is a collection of terms that have semantic relations together. We have applied two different topic modeling methods: LDA and LSA.

LDA is an unsupervised generative probabilistic model based on the BOW (Blei et al. 2003). It generates a distribution of latent topics over each document and each topis is represented by the distribution of words**.** LSA detects semantic relations between terms and documents and employs a singular value decomposition (SVD) to reduce the data dimension and obtain the k best approximation of the original data (vector format).

Although the probabilistic background in LDA is an advantage over LSA and helps LDA work effectively in more complex models (Blei et al. 2003), it has been proved that LSA can discover some topics which LDA cannot find. These two methods can complement each other (Williams and Betak 2018).

2.3.4 Random Forest

Random forest (RF) is a traditional type of supervised learning developed by (Breiman 2001). RF as an ensemble learning combines several decision tree algorithms to generate a more excellent machine learning method. RF benefits many advantages, such as being effective in a large dataset or less prone to overfitting and noise (Rodriguez-Galiano et al. 2012). Moreover, it provides the important feature and their scores in the classification.

2.3.5 K-Nearest Neighbors

K-nearest neighbors (KNN) assign a class to an unseen object based on K-nearest neighbors of the object (Cover and Hart 1967). This method computes the distance between a new data point and all other data points to find the K-nearest neighbors of the new one. There are various methods to calculate distance, including cosine similarity, Manhattan, Euclidean, and Minkowski.

In this study, we used KNN in two different layers. In the first layer, each article is a data point, and the latent topics of articles are the input features. In the second layer, we used semantic scores as the input features.

2.4 Experiment Setting

2.4.1 Dataset

In this study, we collected three types of news articles: fake news, satire, and real news. Fake news has been scraped from websites such as Politicot, Advocates, and Natural News. The first two websites are notorious for publishing fake news during the election of 2016, while Natural News is notorious for publishing scientific fake and conspiracy theories (Zhang et al. 2019). We collected satire news from famous satire websites such as The onion, Beaverton, and Clickhole. Table 2.1 summarizes the dataset we used in this study.

Source	**Articles**	**Label**
New York Times	4862	Real
NBC	2957	Real
Advocate	3460	Fake
Natural News	2396	Fake
Politicot	3056	Fake
The Onion	3396	Satire
Beaverton	1207	Satire
Clickhole	816	Satire

Table 2.1: The distribution of three categories of news and website sources.

2.4.2 Feature vector and detection model

We leveraged the Scikit-learn package from Python to perform the LDA and LSA model on the data. We applied LDA and LSA to BOW and TF-IDF vectors and set both models on 10 topics. We did not use any perplexity or coherence method to evaluate the LDA and LSA. Instead, we evaluated them based on the performance of the detection model explained in the "Sensitivity Analysis For Number of Topics" section (2.5.2).

The results of these two models were the topic probability distribution in each document. Table 2.2 shows an example of topic probability distribution obtained from LDA. We used the results as feature vectors in our multi-class RF detection model.

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
Article 1	0.24	0.12	0.41	0.00	0.00	0.00	0.00	0.10	0.00	0.11
Article 2	0.00	0.00	0.92	0.00	0.00	0.00	0.06	0.00	0.00	0.00
Article 3	0.05	0.60	0.16	0.00	0.03	0.05	0.00	0.03	0.08	0.00
Article 4	0.02	0.29	0.00	0.00	0.00	0.00	0.00	0.00	0.17	0.51
Article 5	0.00	0.00	0.65	0.00	0.24	0.00	0.00	0.00	0.09	0.00

Table 2.2: An example of topic distribution probability in each article

2.4.3 Recommended News

In the second stage of this project, we aimed to make a recommendation based on two categories of features: topic probability distribution and sentiment score. The sentiment score represents the reader's mood, where if a user was reading good news (happy) or bad news (sad), the model continues recommending good or bad news. Regarding the sentiment analysis, we used the Valence Aware Dictionary and Sentiment Reasoner (Vader) package in Python that shows the sentiment of news with a numeric value number between -1 to 1.

To apply KNN, we employed the Scikit-learn library in Python and set the distance metric to Minkowski, a combination of Euclidean distance and Manhattan distance. In the first layer, we set k = 20 to show 20 similar news articles to the last read news article by a user. We filtered out 10 of those news pieces in the second filter and kept 10 articles based on the sentiment score similarities.

2.5 Results and Discussion

2.5.1 Validation Metrics

The performance of the detection model is evaluated by measuring accuracy, sensitivity, precision, and F1-measure. Our model can predict three categories with 85% accuracy, and other metrics are presented in Table 2.3.

Category	Sensitivity	Specificity	Precision
real	0.88	0.94	0.89
Fake	0.84	0.86	0.83
Satire	0.83	0.94	0.82

Table 2.3: Validation metrics

In addition to the metrics mentioned above, we also calculated accuracy based on 10-fold cross-validation in which the training set was divided into 10 sets. Then, the first set was kept for testing, while the remaining 9 sets were used for training. This process was repeated 10 times, and the test set was changed in each run. The accuracy is calculated based on the average accuracy of these 10 runs. The overall accuracy of the 10-fold cross-validation is 0.84.

2.5.2 Sensitivity Analysis for Number of Topics

The number of topics is a hyperparameter in topic modeling and can affect the experiment results. A large number of topics makes the results very specific, and a small number of topics makes them very general (Pavlinek and Podgorelec 2017). There is no universal confirmed formula for determining the number of topics. The optimum number of topics in topic modeling can be a topic of research, like the study done by (Greene et al. 2014). Some metrics, like perplexity proposed by (Blei et al. 2003), are also used to evaluate the topic model. However, in this study, we found the number of topics based on the best balance between the number of topics, computational expense (execution time), and the accuracy score of the detection model. We ran the detection model with feature vectors obtained from the LDA and LSA, setting them on 5, 10, 15, and 20 topics. Then, we checked for any changes in the accuracy. Table 2.4 shows how accuracy changes relative to the different number of topics. The results in Table 2.4 show that the best balance happens when we use 10 topics while 15 and 20 topics increase executive times but do not improve accuracy. Therefore, we set the number of topics equal to 10.

Number of topics	LDA execution time	Accuracy	LSA execution time	Accuracy
5	189.15 s	74%	54.90 S	0.74
10	217.12 s	81%	71.41 S	0.80
15	248.84 S	81%	75.22 S	0.81
20	251.88 S	81%	87.36 S	0.81

Table 2.4: Comparing accuracy and execution time in different number of topics

2.5.3 Feature Importance

Feature importance is one of the advantages of RF and provides the rank of each feature based on how each feature contributes to detection. In this study, we benefit from this property to identify the topics with discriminative attributes that can play a predictor role.

Mangal and Holm (2018) discussed how RF finds feature importance. RF is a collection of trees, and in each tree, one feature node is randomly chosen and replaced by another feature, while other nodes remain unchanged. Then, if the previous feature node was important, the result of the detection would decrease. Therefore, RF provides a feature importance score based on the ability of a feature to discriminate different categories. We used the Scikit-learn package in Python to apply RF. Table 2.5 illustrates the top five most important features. The accuracies in Table 2.5 are cumulative (from the left column to right) and indicate that the top ten important features constitute 83% accuracy out of 85%.

feature	Distribution topic # 8 from LDA	Distribution topic # 1 from LDA	Distribution topic # 3 from LSA	Distribution topic # 5 from LSA	Distribution topic # 3 from LDA
Accuracy	52%	69%	74%	77%	80%

Table 2.5: Ranking of feature importance

To better understand the important features in our dataset, we presented the top 10 terms for each of the 10 topics in LDA and LSA in Tables 2.6 and 2.7, respectively. The bold topics are the 5 top important ones. Although we filter out stop words and the words that appear in more than

95% of the document, other words such as “say” look to have a low semantic contribution. However, further analysis shows that removing “say” from the content resulted in an accuracy reduction to 3%. The results show that topics 1 from LDA and 3 from LSA are more about LGBT, and topics 3 from LDA and 5 from LSA are more about health. The top 10 terms in this topic 8 from LDA do not show the theme behind the topic, though it is one of the 5 important topics.

Topic 1	**Topic 2**	**Topic 3**	**Topic 4**	**Topic** 5	**Topic** 6	**Topic** 7	**Topic** 8	**Topic** 9	**Topic 1**0
court	say	**health**	say	say	trump	north	**time**	say	say
state	like	**study**	city	news	presiden	party	**like**	company	united
law	woman	**people**	police	report	say	election	**go**	year	governm
marriag	man	**find**	people	medium	house	vote	**know**	new	states
LGBT	people	**say**	state	student	Obama	say	**say**	percent	country
gay	time	**food**	home	post	white	Clinton	**people**	million	official
right	year	**drug**	kill	Facebook	administ	voter	**think**	market	military
sex	come	**year**	family	write	republic	campaig	**thing**	pay	china
people	know	**percent**	tell	new	senate	candidate	**get**	business	America

Table 2.6: Top 10 words in LDA

Topic 1	**Topic 2**	**Topic 3**	**Topic 4**	**Topic** 5	**Topic** 6	**Topic** 7	**Topic** 8	**Topic** 9	**Topic 1**0
say	trump	**gay**	marriage	**study**	company	party	north	court	woman
trump	president	**marriage**	court	**health**	party	obama	china	marriage	north
people	donald	**lgbt**	state	**trump**	vote	vote	gay	north	china
president	clinton	**sex**	law	**marriage**	bill	state	country	supreme	court
year	campaign	**court**	government	**researcher**	market	republican	lgbt	couple	marriage
like	obama	**equality**	federal	**disease**	voter	house	united	president	party
new	election	**man**	china	**woman**	election	bill	military	obama	clinton
time	house	**right**	equality	**percent**	republican	voter	south	sex	voter
state	white	**couple**	north	**food**	billion	election	party	judge	percent
know	republican	**woman**	lgbt	**sex**	business	senate	obama	child	candidate

Table 2.7: Top 10 words in LSA

2.5.4 *Implications for research and practice*

In this study, we developed a news recommendation system that recommends news based on latent topics and the sentiment score similarities. In the proposed model, we only recommend news based on its latent topics and sentiment score of the article's content. Future studies can improve this model by adding more data and more features such as user's demographic.

Moreover, social media developers can take advantage of the proposed model to recommend real news to the users and fight the creation of the filter bubble and echo chambers.

2.6 Conclusion

The new generation of digital news provides a wealth of opportunities for people to easily access news from their social media accounts, such as social networks or link-sharing websites. Despite the advantages of digital news, accessing news from social media impose significant threats. On the one hand, on this platform, users easily publish and share news, regardless of whether that news is real or not. Therefore, many users can easily be fooled by fake and satirical news. On the other hand, users might be trapped in a filter bubble and echo chamber created by social media's recommendation systems. The echo chamber and filter bubble increase the likelihood of exposing the user to fake news, in which users are recommended similar news articles based on their past views and regardless of the type of news. To combat fake news dissemination and reduce the likelihood of exposing fake and satirical news we proposed a news recommendation system. For example, if a user reads extreme right new articles, the recommendation system isolates the user by recommending the extreme right articles. The proposed model fights fake news in two out of three stages in the life cycle of news in social media (creation, publication, and propagation): Publication and propagation.

Our proposed system includes two stages. In the first stage, we made a detection model based on an RF method that classifies and detects an article as fake, satire, and real. The features used in this model are the topic probability distribution of documents obtained from LDA and LSA. At the end of the first stage, the model filters out fake and satirical news and stores real news in a separate repository (legitimate repository). In the second stage, we used two filters to recommend news to a user from the legitimate repository. In the first filter, a KNN model looks for the legitimate repository to find 2 N similar news to the user's last read news article based on topic probability distribution. In the second filter, we perform sentiment analysis, and then the KNN model picks N similar news based on sentiment score out of 2 N collected news in the first layer.

The result of this study shows that the proposed model can detect the type of news with 85% accuracy in the publication stage, and the recommendation system reduces the spread of fake news propagation stage by filtering out the fake and satire news and recommending real news.

3 Analyzing the language used in fake news related to COVID-19

3.1 Introduction

Although the term “fake news” started to appear in media and research studies after the 2016 election, the history of fake news originated long before the election, where a broader concept like misinformation has always been an issue, especially during or after a crisis. For example, during the HIV epidemic, misinformation was raised regarding the existence of this virus or remedies (Mian and Khan 2020). Similarly, misinformation caused a lot of anxiety and fear during H1N1(Chew and Eysenbach 2010a), Ebola (Oyeyemi et al. 2014), and Zika (Ghenai and Mejova 2017). For example, one of the misinformation stories regarding Ebola said that staying with an infected person in the same place is enough to be infected (Spinney 2017). Considering the history of misinformation, it is not surprising to face significant amoutns of fake news in a crisis such as the COVID-19 (coronavirus disease) pandemic.

Since the early days of COVID-19, fake news proliferation on social media has become a major concern for public health to the extent that the Director-General of WHO said, “We’re not just fighting an epidemic; we’re fighting an infodemic” (Zarocostas 2020). During the lockdown and social distancing, people spent a lot of time on social media (Laato et al. 2020), and consequently, they are exposed to fake news (Roozenbeek et al. 2020). A study done by British regulator OFCOM shows that almost half of the UK online adults came across fake news about the COVID-19 (Ofcom 2020). Fake news on COVID-19 covers highly diverse contexts such as conspiracy theories about the virus’s origin, remedies context, and political topics. Since people’s behavior is very effective in managing pandemics, fake news about the COVID-19 played a negative role in managing this crisis (Pennycook et al. 2020a). Even if some fake news stories seem absurd, they still can have the potential to affect and endanger people's life. For example, a conspiracy theory claimed a 5G mobile phone signal can decrease the immune system and contribute to spreading the virus. This fake story resulted in the burning of 5G towers in the UK (Ahmed et al. 2020). Fake news related to treatment is even more disruptive where it can work against health providers' advice with wrong information (Mian and Khan 2020). For example, some fake remedies suggest that vitamin C, salty water, and drinking bleach can treat COVID-19 (Mian and Khan 2020). In Nigeria, several chloroquine poisonings were reported after unverified news regarding the effectiveness of chloroquine in the COVID-19 cure (Busari; and Adebayo 2020).

In response to the full-fledged concern of fake news related to COVID-19, many researchers and practitioners have started a battle against it. The World Health Organization (WHO) has developed a "Mythbuster" website in which they provide a list of detected fake news (Organization 2020). Furthermore, social media like Facebook and Instagram developed an algorithm to detect false claims. Similarly, Twitter deployed an algorithm to show the only reliable source to a user who searches on Twitter (Marr 2020). Furthermore, information system researchers entered the battle to fight against fake news on COVID-19 (Pan and Zhang 2020).

Google trends show the high search frequency of the keywords related to fake news and the coronavirus in the three months between January and May 2020 (Shahi and Nandini 2020). Integration of fake news and COVID-19 also attracted much attention from researchers, evident by the fact that the number of studies about fake news on COVID-19 is increasing, however, due to the novelty and complexity of COVID-19 fake news, there is a lot of room left for research in this area.

In this study, we made a comparative study to investigate and compare COVID-19 real and fake news articles on websites. To this end, we raised different hypotheses based on multiple deception theories and employed NLP and partial least squares structural equation modeling (PLS-SEM) techniques to test those hypotheses in the context of fake news on COVID-19.

The remainder of this contribution is structured as follows. The next section (3.2) presents the related work on COVID-19 fake news. In section 3.3, we explain the theoretical background that supports this study. Section 3.4 illustrates the methods used in this study. The results are described in section 3.5. Finally, the discussion and conclusions are presented in sections 3.6 and 3.7, respectively.

3.2 Related Work

Existing studies on combating fake news related to COVID-19 can be broadly categorized into three groups: (1) Creating a prediction model where the model can discriminate fake news from real news, (2) Analyzing the motivation behind sharing fake news, and (3) Exploratory content analyses.

3.2.1 Detection model

Detecting fake news, in general, highly depend on the NLP methods employed. Dharawat et al. (2020) suggest a model that detects and classifies a tweet into five categories based on the seriousness of misinformation. In another study, Daley (2020) developed a detection model that detects fake news related to COVID-19 based on the features obtained from the headlines. Hossain et al. (2020) made a model that retrieves misconceptions related to a tweet and identifies whether the tweet supports misconception or not. Serrano et al. (2020) employed users' comments as features to identify COVID-19 misinformation videos on Youtube. Al-Rakhami and Al-Amri

(2020) extracted features from tweets and user profiles to detect whether a tweet is credible or not. Khanday et al. (2020) developed a model that detects propaganda about COVID-19 from Twitter.

3.2.2 Exploratory Content Analyses

Several studies regarding COVID-19 were conducted based on exploratory content analysis such as topical and sentiment analysis, which can be applied in news articles and communication posted on social media. Exploratory content analysis provides insight into understanding behavior, magnitude, themes, narrative, public opinion, and sentiment associated with COVID-19. Although numerous studies perform exploratory analysis on broad-ranging COVID-19 information, such as Sha et al. (2020), we narrow our work to the cover articles that included misinformation in their studies. Zhu et al. (2020) applied LDA on textual data obtained from Weibo during the pandemic and obtained 8 topics, including "origin", "host", "organization", "quarantine measures", "role models", "education", "economic", and "rumor". They reported that the highest discussion rate belongs to the origin of the virus while rumors and fake news has the lowest discussion rate. Similarly, Singh et al. (2020) found that less than 0.6% of their sample of tweets discussed myths associated with COVID-19. Mutanga and Abayomi (2020) used LDA to extract topics from COVID-19 related tweets. They found that conspiracy theories and fake news topics are among the most frequently used topics. Subsequently, they provide a list of wrong perceptions that can assist the government in addressing citizens' concerns. In another study, Kouzy et al. (2020) investigated 673 tweets based on 14 hashtags and keywords associated with COVID-19 and found around 42% of tweets included misinformation and unverified information. McQuillan et al. (2020) applied network mapping, LDA, and divergence analysis on Twitter data to investigate the behavior of COVID-19 misinformation networks.

3.2.3 The motivation behind sharing fake news

An important step in developing an intervention against misinformation is to understand why people believe it and contribute to spreading it (Laato et al. 2020, Pennycook et al. 2020a). A study by Pennycook et al. (2020a) shows that encouraging people to think about the accuracy of information increases their power to discriminate truth. In another study, Islam et al. (2020) applied PLS-SEM and neural network techniques on data gathered from adults in Bangladesh. They found that people driven by self-promotion and entertainment or people who suffer from a lack of self-regulation are less careful to publish misinformation regarding COVID-19. Similarly, Apuke and Omar (2020) applied PLS-SEM on data collected from Facebook and WhatsApp in Nigeria and found that altruism, instant news sharing, socialization and self-promotion are significant predictors for fake news sharing behavior, while entertainment is insignificant.

3.3 Theoretical foundation

Inspired by the Undeutsch hypothesis (Undeutsch 1967), which hypothesizes that a true statement’s quality and writing style is different from a fake one, we adopted deceptions theories. We aim to study the possibility of leveraging automating linguistic cues in detecting differences between fake news content and real news content.

3.3.1 Reality Monitoring (RM)

RM was developed by Johnson and Raye (1981) based on the assumption that external memories (externally-derived information) are distinguishable from internal memories (self-generated information) (Masip et al. 2005). Based on RM, describing a memory involves four types of information: contextual such as space and time, sensory or visual details such as colors, semantic information, and cognitive operations (Johnson and Raye 1981). It is assumed that an event based on real experience involves more spatial, temporal, sensory, and semantic information, while an event based on fake experience is much more elaborate (Oberlader et al. 2016).

3.3.2 Four-Factor Theory (FFT)

FFT proposed by Zuckerman et al. (1981) discussed that four factors play a role in revealing cues in deceptive communication. These include (1) behavior control: deceivers' strategies to be as persuasive as possible result in excessive attempts to control their behavior; (2) arousal: fear of being caught might result in anxiety and other psychological arousals; (3) cognitive effort: deceivers think a lot to design details for their lies and conceal inconsistencies in their stories, and; (4) The affective approach: deceivers show negative affect as they feel guilty.

3.3.3 Interpersonal Deception Theory (IDT)

IDT was proposed by (Buller and Burgoon 1996) to study deception in the context of interpersonal communication, where deception is an interactive process between senders and receivers. According to IDT, deception is a goal-oriented behavior in which deceivers try to minimize their responsibility to alleviate unpleasant consequence in case their deceptions are discovered. IDT posits that the deceiver's communication involves both strategic and non-strategic behavior. Strategic behavior refers to the intentional strategy that the deceivers use to deceive receivers and achieve their goals. These strategies can pass through modifying and manipulating messages' completeness, accuracy, and relevance. Non-strategic behavior results in an unintentional behavioral leakage, and deceivers give away cues through reflecting perceptual, cognitive, and emotional behavior (Buller and Burgoon 1996).

3.3.4 Information Manipulation Theory (IMT)

IMT (McCornack et al. 1992) is grounded in four conversational principles: a true message is informative, truthful, relevant, and clear (Grice 1989). On the other hand, a deceiver (1) manipulates the amount of information in a message (Quantity), (2) reduces the quality of information regarding the veracity of the message, (3) reduces or omits relevant information of the message toward the topic of conversation, and (4) makes the message vague and unclear.

3.3.5 Management Obfuscation Hypothesis (MOH)

Unlike the other theories based on communication principles, this hypothesis is based on the assumption that managers conceal their adverse information in their annual reports by increasing obfuscation, especially when the firm's earnings are low (Li 2008).

3.4 Research Model and Hypothesis

We raised different hypotheses to formulate our research questions, study fake news related to COVID-19 in news articles, and compare them to real news articles. Inspired by the deception theories mentioned above, First, we form different hidden strategies that might be used in building fake news. Second, we found possible linguistic cues that help reveal those strategies. Finally, we study the hypothesis in the context of news articles related to COVID-19 to test the strategy differences used in fake news and real news.

3.4.1 Uncertainty

According to IDT and IMT, deceivers use uncertain language. Uncertain language refers to using evasive and vague language (Zhou et al. 2004a) and giving irrelevant information about the topic (Fuller et al. 2013). To measure uncertainty, we utilized linguistic cues, including hedge words (tentative language), subjective vocabulary, and modal verbs. Hedge (tentative) words (such as guess, perhaps, doubt) are the words that reflect possibility and collegiality (Hyland 1998) and indicate the author is uncertain about the topic (Tausczik and Pennebaker 2010). Prior studies discussed that uncertainty could be found in the subjective language (Rubin et al. 2006), where the subjective language is positively associated with biased words (Zhou et al. 2020a). Modality shows commitment and certainty toward a claim (Mitra et al. 2017). Modal verbs such as (can, may, could) are auxiliary verbs that are followed by a verb of prediction (Zhou et al. 2003b). We excluded 'will' and 'must' from the list of modal verbs since they show a low level of uncertainty. The COVID-19 pandemic involves a lot of uncertainty and lack of information which enlarges the stream of infodemic (Lovari 2020). In such a situation, fake news writers take advantage to increase uncertainty in their writing without being caught, while real news writers try to fight and decrease uncertainty. Moreover, uncertainty is considered one of the linguists' cues that is unlikely to be found in news articles (Zhou and Zafarani 2018b). Nevertheless, we assume that, fake news associated with COVID-19 have more uncertainty than real news. Therefore, we draw the first hypothesis:

H1*:* Uncertainty in fake news related to COVID-19 is significantly greater than real news.

3.4.2 Specificity

Based on IDT and RM, deceivers are less specific toward an event. Deceivers' information is not based on real experience, and therefore they cannot provide perceptual and detailed information. (Zhou et al. 2003b). The primary linguistic cues that can measure specificity in a text document include a sensory ratio (such as seeing, feeling, and hearing), space ratio, and time ratio (Zhou et al. 2003b). Moreover, Hauch et al. (2015) hypothesized that deceivers use fewer descriptive words. We included quantifier (each, few), prepositions (about, for), and numbers (one, two) as descriptive words in this study. In another study, Addawood et al. (2019) assumed that deceivers employ fewer discourse markers to be less specific about their fake story. Discourse markers include exclusion words (such as: but and except), conjunctions (such as: also and although), and negations. Therefore, we used the aformentioned cues to measure specificity in the fake news related COVID-19 and raise the following hypothesis:

H2: Specificity in fake news related to COVID-19 is significantly less than real news.

3.4.3 Immediacy

Immediacy is one of the assumptions in IDT, and it is based on the premise that the deceivers try to stay away from accepting the responsibilities of their statements. Therefore, deceivers use words with the low immediacy and high non-immediacy. For example, they prefer not to use first-person references like "we" and "I" that have higher immediacy than second or third references like "you" and "they." Besides, passive voice is another trick that the deceivers might use to reduce and push away the responsibilities (Zhou et al. 2003b, Zhou et al. 2004a). Finally, indefinite articles (such as "a" and "the") increase the distance of the deceiver to the event and refer to a general concept (Addawood et al. 2019). Therefore we test the hypothesis:

H3: Non-Immediacy in fake news related to COVID-19 is significantly higher than real news.

3.4.4 Information Quantity

One of the assumptions in IMT is that deceivers manipulate the amount of information, and their messages are not as informative as truth-teller. Telling a fake story demands a high cognitive load, while deceivers may not be able to provide information as much as a truth-teller (Hauch et al. 2015). Previous studies, such as (Burgoon et al. 2003) proved that deceivers' messages are shorter than the truth-tellers. Zhou et al. (2004b) used the number of words, adjectives, and adverbs as linguistic cues to measure quantity. We included the number of sentences, adjectives, and adverbs in our analysis. Therefore, we hypothesize:

H4: information quantity in fake news related to COVID-19 is significantly less than real news.

3.4.5 Affect

According to FFT and IDT, negative emotions such as anxiety and fear can be found in the deceiver's behaviour. Zhou et al. (2004a) studied e-mail messages and found that the deceiver tends to use more affective messages, including both positive and negative emotions. Other studies, such as Brady et al. (2020) and Vosoughi et al. (2018), discussed that affective content attracts more attention, and that is one reason that fake news spreads much faster than real news. The result regarding this assumption is mixed and shows that depending on the context and situation, deceivers might use either more negative or positive emotions or both. For example, on abortion, the deceivers prefer to use more negative emotional words (Newman et al. 2003). Supprisingly, Toma and Hancock (2010) showed that deceivers use fewer negative emotions in online dating. As COVID-19 creates a lot of anxiety and stress in the population, we expect to have a more negative affect on fake news related to COVID-19 than real news. We measure the negative affect by counting the average number of words related to death, risk, anxiety, anger, and sadness. Therefore, we hypothesize that:

H5: negative affect on Fake news related to COVID-19 is significantly higher than real news.

3.4.6 Cognitive process

Cognitive load is one of the assumptions of RM, which assumes that the deceivers have to bear a higher cognitive load to make up a memory about an event. Also, IDT emphasizes that increased cognitive load makes deceivers deviate from their normal. The deceivers have to think and keep track of their previous statements (Hancock et al. 2007). The results and linguistic cues are different in different studies. Ho et al. (2016) studied the cognitive process in spontaneous online communication. They found that words associated with insight were used significantly higher by deceivers than truth-tellers, while they found more words related to insight in true accounts. We also used words related to insight (think, guess) to measure cognitive process and raised a hypothesis:

H6: cognitive process in fake news related to COVID-19 is significantly higher than in real news.

3.4.7 Informality

Siering et al. (2016) utilized IDT to hypothesize informality. Prior studies proved that informality could be considered one of the linguistic cues in deception detection. For example, Zhou et al. (2004a) and Zhou et al. (2003a) studied deception in text-based electronic communication and E-mail context, respectively. Both studies proved that deceivers use more informal language. However, in the context of news articles, informality might not be a cue (Zhou and Zafarani 2018b). Some studies measure informality by counting the proportion of misspelling words, but in this study, we count the proportion of words associated with swearing, netspeak (such as: plz, thanx, prob), assent (such as absolutely, cool), non-fluencies (such as ah, oh) and filler (anyway, blah). Therefore, we hypothesize:

H7: informality in fake news related to COVID-19 is significantly higher than real news.

3.4.8 Complexity

Complexity is one of the premises of IDT and IMT. Deceivers consume higher cognitive resources, forcing them to use a simpler story (Newman et al. 2003). However, on the other hand, according to the management obfuscation hypothesis, more obfuscation and less readability help deceivers hide the fraud. Therefore, we expect that deceivers use more complexity in vocabulary level to make their statements less readable. To cover both assumptions of complexity, we categorized and measured complexity on two levels: sentence and vocabulary.

Regarding sentence complexity, as Newman et al. (2003) explained, deceivers use a simple sentence like “ I walked home”, while truth-tellers can assess their statements with causation words (e.g., “Usually I take the bus, but it was such a nice day”). In another study, Hancock et al. (2007) hypothesized that causal terms and phrases put a deceiver at risk of detection. In addition, it is expected that deceivers keep their sentences simple and use fewer words and punctuation in each sentence than a truth-teller.

To measure complexity in the vocabulary level, we used two features. The first feature is readability which is in line with a study done by (Horne and Adali 2017), and the second feature is the average number of syllables per word. In contrast to complexity in the sentence level, we expect that deceivers use more complexity at the vocabulary level to make their statements less readable. According to the management obfuscation hypothesis, more obfuscation and less readability help deceivers hide the fraud. We used the Flesh-Kincaid grade level to measure readability. Therefore, we assume that fake news writers on the COVID-19 subject use more straightforward sentences while making their statements less readable. We set two different hypotheses:

H8: Complexity in the sentence level in fake news related to COVID-19 is significantly less than real news.

H9: Complexity in vocabulary level in fake news related to COVID-19 is significantly higher than real news.

For the rest of this study, we used complexity 1 to refer to complexity at the sentence-level and complexity 2 to refer to complexity in vocabulary.

3.4.9 Diversity

Diversity is another premise of RM, where it assumes that deceivers may not be able to use rich lexical diversity in describing a fake event. Previous studies such as (Zhou et al. 2004a, Zhou et al. 2003a, Zhou et al. 2003b) proved a significant difference between truth-tellers and deceivers regarding lexical diversity. We measure diversity on three levels: lexical, functional, and dictionary words (ordinary words). First, we counted the average number of unique words to measure lexical diversity (Zhou et al. 2003b). Second, we computed the average number of function words, which is a redundancy measurement (Zhou et al. 2004a). Third, we calculate the average number of dictionary words which is the measurement to evaluate normal language (Peden and Carroll 2008). We used the LIWC dictionary for the third approach. Finally, we expect that deceivers use a smaller number of lexical diversity, function words, and everyday words than truth-teller. Therefore, we test the hypothesis:

H10: Diversity in fake news related to COVID-19 is significantly less than real news.

3.4.10 Expressivity

Expressivity is a premise suggested by IDT, which refers to vivid and emotive language (Fuller et al. 2013). A deceiver employs more emotional language and increases expressivity. Zhou et al. (2004a) discussed that in the context of e-mail messages, deceivers use more expressive language. In another study, Humpherys et al. (2011) found no significant difference in expressivity between fraudulent and non-fraudulent financial disclosures. We measure the expressivity by computing the proportion of adjectives and adverbs to the number of nouns and verbs, as suggested by other studies such as Siering et al. (2016) and Zhou et al. (2004a). Therefore, we raise the hypothesis:

H11: Expressivity in fake news related to COVID-19 is significantly greater than real news.

3.5 Research methodology

In this study, we performed four steps to complete our comparative analysis. In the first step, we collected two categories of data: fake news and real news related to COVID-19. In the second step, we applied text pre-processing techniques; in the third step, we extracted the linguistics feature, and finally, we used an analytical model. Moreover, we employed MANOVA and a prediction model to support our method.

3.5.1 Data collection

To study and compare the differences between fake and real news related to COVID-19, we collected and created a separate repository for each category of data. To create a fake news repository, we developed a web crawler to collect data from a list of unreliable sources used by Gupta et al. (2022). Table 3.1 shows the distribution of data, data source, and their categories.

Source	**No. Articles**	**Label**
Reuters	1585	real
NYPost	2490	real
NYTimes	160	real
BBC	151	real
Time	112	real
The Jim Bakker Show	1214	Fake
Newspunch	691	Fake
Clashdaily	402	Fake
The Epoch Times	1514	Fake
Infowars	551	Fake
Red State	68	Fake
Dcgazette	29	Fake
Intellihub	7	Fake

Table 3.1: Data distribution

3.5.2 *Text Pre-processing*

We applied NLP techniques to clean and prepare data for the next steps. First, we tokenized all the raw texts into the unigram token, where each token is a word. Then, we converted all words into lower case and removed the noise. Finally, we applied lemmatization techniques, as some features are based on the regular form of the words. All the steps in text-preprocessing were executed in the Spacy package in Python (Honnibal and Montani 2017).

3.5.3 *Feature Extraction*

We converted all raw text to numeric format and created a wide range of features where each feature is considered an observed variable (indicators) and represents a linguistics cue. The techniques and tools employed here include POS tagging, LIWC (Pennebaker et al. 2015), and bias lexicon corpus (Recasens et al. 2013).

The features used in this study can be categorized into two categories, latent features (exogenous latent variables) and indicator features (endogenous indicator variables). Latent features cannot be measured directly and need to be measured indirectly by using indicator features. In our case, the latent features are the hypotheses, and indicators are the linguistic cues. We applied a partial PLS-SEM technique to achieve this goal.

3.5.4 *PLS-SEM*

PLS-SEM is a multivariate statistical technique that leverages multiple regression analysis and executes a path analysis to estimate the relationship between all variables (Ullman and Bentler 2003). One of the main advantages of PLS-SEM is the ability to interpret the relationship between latent factors, which can be measured by indicator factors (Jaccard and Wan 1995, Wolf and Seebauer 2014). The PLS-SEM includes two subset models: the measurement model and the structural model. The measurement model is the part of the model used to verify latent variables and how they are constructed by observed variables (indicators). The structural model analyzes the relationships among latent variables using multiple regression analysis and path analysis (Qureshi and Kang 2015). Figure 3.1 depicts the relationships among the latent features and target (Fake) in the structural model and illustrates the hypothesis we need to test.

Since our data has a binary target variable (that indicates if an article is a fake news article or a real article), we employ a two-stage approach suggested by (Bodoff and Ho 2016) which combines PLS-SEM with Logistic regression. In the first stage, we applied PLS-SEM by utilizing smart pls 3.0 software (Ringle et al. 2015) to obtain the score of the latent variables. In the second stage, we used SPSS to apply logistic regression to latent scores.

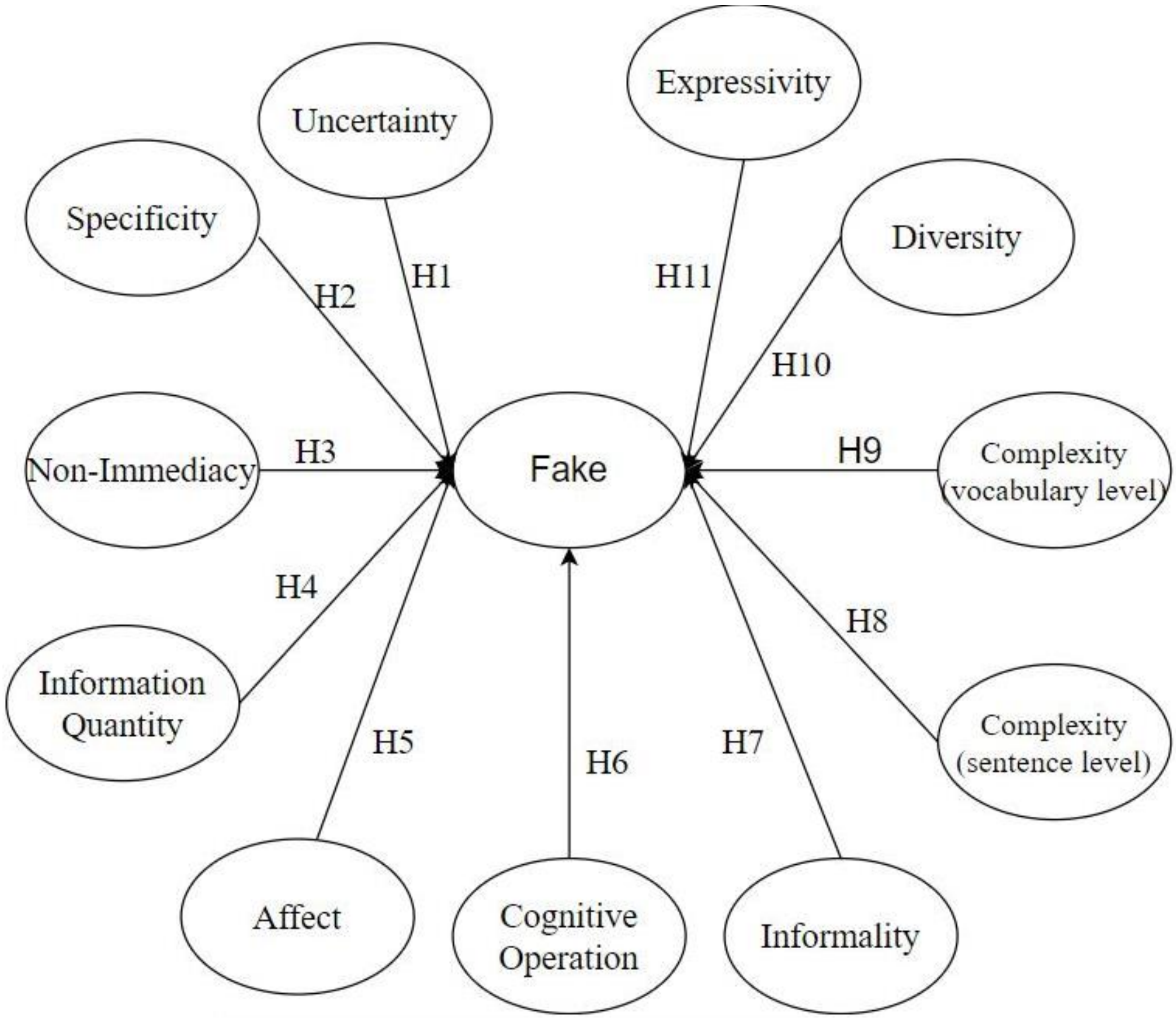

Figure 3.1: Structural model and hypothesis

We coded fake news articles related to COVID-19 as 1 and real news as 0, and then we made a formative construct in both the measurement model and structural model. Table 3.2 shows all latent variables, their indicators in the measurement model, and the tools used to extract them.

Latent Variables	Indicators	Tools
Uncertainty	The proportion of the bias terms	(Recasens et al. 2013)
	The proportion of modal verbs	Self-implemented
	The proportion of tenant terms	LIWC
Informality	The proportion of assent terms	LIWC
	The proportion of filler terms	LIWC
	The proportion of netspeak	LIWC
	The proportion of swear terms	LIWC
Cognitive operations	The proportion of insight terms	LIWC
Non-Immediacy	The proportion of articles	LIWC
	Passive voice	LIWC
	The proportion of singular third person	LIWC
	The proportion of plural third person	LIWC
	The proportion of second person	LIWC
Diversity	The proportion of dictionary terms	LIWC
	The proportion of function terms	LIWC
	The proportion of unique terms	Self-implemented
Quantity	Number of adjectives	LIWC
	Number of adverbs	LIWC
	Number of sentences	LIWC
Complexity sentence level	The proportion of punctuation	LIWC
	The average number of words per sentence	LIWC
	The proportion of causation terms	LIWC
Complexity vocabulary level	Average of word length	Self-implemented
	Readability (Flesch–Kincaid)	Self-implemented
Specificity	The proportion of feeling terms	LIWC
	The proportion of seeing terms	LIWC
	The proportion of hearing terms	LIWC
	The proportion of space terms	LIWC
	The proportion of time terms	LIWC
	The proportion of quantifier	LIWC
	The proportion of numbers	LIWC
	The proportion of negation terms	LIWC
	The proportion of conjunction terms	LIWC
	The proportion of quantifier	LIWC
	The proportion of preposition	LIWC

	The proportion of motion terms	LIWC
	The proportion of exclusion terms	LIWC
Negative affect	The proportion of sadness terms	LIWC
	The proportion of death terms	LIWC
	The proportion of risk terms	LIWC
	The proportion of anger terms	LIWC
	The proportion of anxiety terms	LIWC
Expressivity	The proportion of adjectives and adverbs to the number of nouns and verbs	Self-implemented

Table 3.2: Latent and indicator features and the tools used to extract the features

3.6 Result and Finding

3.6.1 Measurement model

As suggested by (MacKenzie et al. 2005, Diamantopoulos and Winklhofer 2001a), We employ two approaches to check the quality of our formative measurement model. First, we studied the multicollinearity among the indicators through the variance inflation factor (VIF) technique. We considered 10 as a cut-off threshold for multicollinearity as suggested by Diamantopoulos and Winklhofer (2001b). We found that all VIFs are between 1 and 2.28 except for the indicators related to Information Quantity which show VIF between 4.26 and 5.6, though they are still less than the threshold. Therefore, we do not exclude any variable in this step. Second, we tested indicator weights to check which indicators have roles in constructing the latent features (Hair Jr et al. 2016). We applied the bootstrapping resampling method (5000 samples) to test the significance level of each indicator. Table 3.3 shows the result obtained from bootstrapping for the relationship between indicator features and their latent features. The results suggest which indicators are significant and contribute to constructing their latent features. We found the causation feature is not significant in constructing complexity 1, number of unique words feature is not significant in constructing diversity, second person is not significant in constructing immediacy, assent and filler are not significant in constructing informality and negation, quantifier and space are not significant in constructing specificity. Other than the mentioned variables, all other variables are significant in constructing their latent variables. The significance of variables also are shown in table 3.3 where in the significant columns “Yes” indicate the variable is significant.

In the next step, we provide which latent variables are significant. The results of the next step can lead us to answer the raised hypotheses.

Latent Variable	Indicators	T Statistics	P Values	Significant
Cognitive	insight			
Complexity 1 (sentence and lexical level)	words per sentence	10.347	0.000	Yes
	causation	0.035	0.972	No
	AllPunc	28.186	0.000	Yes
complexity2 (vocabulary level)	avg_wordLength	10.804	0.000	Yes
	gunningfog_score	5.037	0.000	Yes
Diversity	dictionary words	5.996	0.000	Yes
	function words	5.485	0.000	Yes
	uniqe_words	0.254	0.800	No
Expressivity				
Immediacy	article	6.759	0.000	Yes
	passive	5.621	0.000	Yes
	singular third person	3.915	0.000	Yes
	plural third person	2.453	0.014	Yes
	second person	1.756	0.079	No
Informality	assent	1.816	0.069	No
	filler	1.924	0.054	No
	netspeak	7.116	0.000	Yes
	swear	9.841	0.000	Yes
Negative affect	risk	5.066	0.000	Yes
	sadness	6.55	0.000	Yes
	anger	8.661	0.000	Yes
	anxiety	4.411	0.000	Yes
	death	12.901	0.000	Yes
Quantity	adjective_quantity	6.527	0.000	Yes
	adverb_quantity	7.186	0.000	Yes
	sent_count	2.23	0.026	Yes
Specificity	conjuction	4.809	0.000	Yes
	exclusive	2.156	0.031	Yes
	feeling	7.062	0.000	Yes
	hearing	10.536	0.000	Yes
	motion	2.587	0.010	Yes
	negation	0.212	0.832	No
	number	12.718	0.000	Yes
	preposition	4.975	0.000	Yes

	quantifier	1.212	0.225	No
	seeing	5.328	0.000	Yes
	space	0.581	0.561	No
	time	24.284	0.000	Yes
Uncertainty	bias terms	64.691	0.000	Yes
	modalverb_ratio	5.327	0.000	Yes
	tentat	5.932	0.000	Yes

Table 3.3: PLS result for the relationship between indicators and their latent variables. Significant at 0.05 level

3.6.2 *Hypothesis Testing*

Table 3.4 present the result of path modeling in the structure model. The result indicates that all latent features are important at a significant level of 0.05, except immediacy. Our model can explain 25% of the variance in fake news versus real news related to COVID-19.

Latent variables	**Coefficient**	**P-values**
Cognitive operations	0.104	0.000
Complexity 1	-0.310	0.000
Diversity	-0.410	0.000
Immediacy	-0.054	0.032
Informality	0.335	0.000
Expressivity	0.121	0.000
Negative affect	0.196	0.000
Quantity	-0.104	0.000
Specificity	-0.454	0.000
Uncertainty	0.422	0.000
Complexity 2	0.267	0.000

Table 3.4: Path coefficient of the structural model

3.7 Discussion

We relied on different psychological theories to explore the differences in the writing style of a fake news writer and a real news writer. To this end, we utilized PLS-SEM and logistics regression. In this section, we summarized our findings and supported them by applying

MANOVA and a detection model. Then we discuss limitations of the study, theoretical and practical implications, and future studies.

3.7.1 Key findings and limitations

The result of this study indicates that specificity is the most powerful strategies used by COVID-19 related fake news writers and is followed by uncertainty and diversity. In specificity, words related to time, numbers, and hearing are the top three important indicators indicating that fake news writers were not able to support their articles with these details as much as real news writers do. In uncertainty, words related to subjectivity (bias terms) are the most significant indicators, proving that the language used in fake news related to COVID-19 is much more subjective than real news. On the other hand, non-immediacy, information quantity and the cognitive process have the weakest impact in distinguishing between fake news and real news in the subject of COVID-19. Complexity 1 and informality are the fourth and fifth important factors in identifying fake news. The average number of punctuations is significantly more in real news and is the most important indicator in complexity 1. Fake news writers use more informal language than real news writers. The most significant indicators of this hypothesis are the proportion of swear words and netspeak, which are significantly higher in fake news articles. Negative affect and complexity 2 are the following two important latent features. Interestingly, the words related to death and anger appeared in fake news significantly more than in real news and are the most important indicators of negative affect.

Further, we used MANOVA to compare the differences between the two groups (fake news vs. real news) based on all latent features. The results of the multivariate test such as Pillai's Trace, Wilks' Lambda, Hotelling's Trace, and Roy's Largest Root prove that there are significant differences between fake news and real news when considering all 11 latent variables at a 0.001

significant level. In addition, the post hoc analysis (Games-Howell test) show all hypotheses are significant. Table 3.5 depicts the result of the multivariate test obtained by SPSS.

Effect	**Test**	**Value**	**F**	**Sig.**
Fake news vs real news	Pillai's Trace	0.18	178.451b	0.000
	Wilks' Lambda	0.82	178.451b	0.000
	Hotelling's Trace	0.219	178.451b	0.000
	Roy's Largest Root	0.219	178.451b	0.000

Table 3.5: Multivariate Test

Besides, we applied XGBoost as a detection model in which our features are the latent scores obtained from the Smartpls. This model helps us test the power of our obtained latent features in detecting fake news related to COVID-19. We split the data into 80% and 20% in which we train the model on the 80% section and test it on the 20% section. The result of the prediction model indicates that the score of latent features can detect fake news with 72% accuracy. Table 3.6 shows the results of the prediction model and confusion matrix.

Label	**Precision**	**Recall**
Real news	0.72	0.70
Fake news	0.71	0.73

Table 3.6: Result of the prediction model

	Actual real	Actual fake
Predicted real news	634	266
Predicted fake news	243	652

Table 3.7: Confusion matrix

3.7.2 Theoretical and Practical Implications

This study compares fake news and real news by raising different hypotheses. Although detecting fake news in the context of news articles is very hard for a human, we believe that the results of this study can improve society's perceptiveness and reasoning power in detecting fake news. For example, based on the result of this study, if an article related to COVID-19 looks very subjective in language, the reader should doubt the article's trustworthiness. Also, our findings can help researchers improve their detection models in two ways. First, in this study, each indicator is a linguist and textual feature which can be used as an input feature in a detection model. Second, we provided which features significantly distinguish fake news from real news. Therefore, it provides insight for the developers of detection models on which features they can focus.

3.8 Conclusion

This study investigates and compares the writing style and strategy used in fake news related to COVID-19 versus real news. To this end, we utilized different deception theories to find the possible strategies a deceiver might use to deceive others. Inspired by deception theories, we raised 11 hypotheses to test the strategies in the context of fake news articles related to COVID-19. Each hypothesis was tested based on a latent variable (feature) and each latent variable was measured by different indicators represented by linguistics features. The methods used in this study were PLS-SEM and logistic regression, where PLS-SEM finds the scores of latent features, and Logistic regression finds the patch coefficient between latent features and the target variable. The results prove that fake news writers on the subject of COVID-19 mostly adopted the same strategies as those deception strategies claim. For example, this study indicates that fake news related to COVID-19 has more uncertainty, less complexity in their sentence, more complexity in vocabulary level, more negative affect, less specificity, less information quantity, less diversity, more expressivity words, and more cognitive process. The only hypothesis that was not supported by this study was immediacy which our result shows no difference between fake and real news. Further, we elaborated and supported our methods by applying MANOVA to two categories of news articles that we had (fake news vs. real news) and creating a detection model in which latent variables were used as the features of the model. The result of MANOVA confirms that there is a significant difference between fake news and real news related to COVID-19, considering all latent variables jointly. The prediction model show that our model can detect fake news with 72% accuracy.

4 Analyzing the characteristics of Facebook Posts about COVID-19 Vaccine

4.1 Introduction

In the 20th century, vaccines played a key role in reducing childhood morbidity and disease outbreaks (Kestenbaum and Feemster 2015). In many diseases, vaccines have been proven to protect individuals against viruses and improve public health (Lahariya and Care 2016). In addition, when the number of vaccinated individuals goes beyond a specific threshold, it can result in herd immunity in the population, which can stop the spread of the virus or decline the speed of spread (Fine et al. 2011). In the current COVID-19 pandemic, developing and utilizing an effective vaccine is the best possible approach to control the spread of the novel coronavirus (Caserotti et al. 2021). Experts suggest that the threshold of herd immunity in the novel coronavirus is between 70% to 85% (Maxouris and Vera 2021). Scientists started to develop an effective vaccine to fight this virus, and finally, different vaccines were built with efficacy between 70% to 95% (Dhama et al. 2021). While developing such vaccines creates hope to defeat the pandemic, there is growing concern regarding people who are against vaccines or have a skeptical view about the vaccines. Many people refuse or delay accepting the vaccine, despite the availability and high efficacy of the COVID-19 vaccines (Dhama et al. 2021). This behavior is called vaccine hesitancy and was ranked among the top ten threats to global health in 2019 by World Health Organization (WHO) (Organization 2019). A recent survey suggests that nearly 23% of U.S. adults are unwilling to receive the vaccine (Trujillo and Motta 2020). Similarly, another survey in Canada shows that only 48% of Canadians are keen to receive the vaccine immediately, 31% prefer to wait a while, and the remaining proportion is either unsure or against the vaccine (AGNUS 2020). Even in the United States, where COVID-19 vaccinations began on December 14, 2020, after more than 15 months (April 2022), around 66% were fully vaccinated, with this number being around 50% for some states. Lack of a wide acceptance made Dr. Anthony Fauci zsaid that “not enough people have been vaccinated” (Goodman 2022). Many reasons are counted for vaccine hesitancy, and they can be categorized into four broad categories including religious reasons, safety concerns, personal beliefs, or philosophical reasons (McKee et al. 2016). In addition to the mentioned categories, there is another group that can be classified under vaccine hesitancy called anti-vax attitudes. Anti-vax groups have negative attitudes toward vaccines (Gravelle et al. 2022) and many times they share misleading information to persuade others. Although vaccine hesitancy is not a new phenomenon, social media intensified society's concerns and accelerated the spread of vaccine hesitancy views (Karafillakis et al. 2021). However, social media could be leveraged to alleviate vaccine concerns and encourage vaccination.

The growth of social media has considerably changed the way people interact and receive information. Social media allows the user to share their opinion and news of interest in different formats such as text, image, and video (Jawad et al. 2021). This technology provides the users with services so they can interact with their friends and family or make a social group with other people who have similar opinions. Creating communication and sharing information on social media during a crisis is not limited to normal users. Many agencies, politicians, health authorities, government organizations, and celebrities also have official social media pages which help them communicate with the public, where they can understand public opinion and sentiment about an action (Raamkumar et al. 2020). Moreover, social media has been proven to be an effective tool in shaping public attitudes. For example, in the marketing domain, social media influencers affect the purchase intention of potential customers of a brand (Lim et al. 2017). Several studies have investigated how social media can affect vaccination decisions in the individual (Moran et al. 2016, Basch and MacLean 2019, Piltch-Loeb et al. 2021). For example, national experts found that social media can increase the vaccination rate with the Human Papillomavirus Vaccine (Reiter et al. 2018). Vaccination hesitancy groups, along with other online platforms such as blogs and forums also leveraged social media to disseminate their attitude and anti-vaccine groups share misinformation such as vaccines causing autism through social media (Leask et al. 2014, Piltch-Loeb et al. 2021, Moran et al. 2016). Exposure to the opinion or message that magnifies the risk of vaccines and undermines the perception of vaccine benefits might increase vaccine hesitancy among individuals (Betsch et al. 2010). (Leask et al. 2014, Piltch-Loeb et al. 2021, Moran et al. 2016). In other words, sharing vaccine information or opinion can affect an individual's decision about whether or not to accept vaccination (Piltch-Loeb et al. 2021).

Misinformation is also involved in vaccine hesitancy. It has been discussed that there is an association between conspiracy beliefs and vaccine hesitancy (Salali and Uysal 2020, Sallam et al. 2021). Generally, there are similarities between vaccine hesitancy messages and misinformation in terms of consumption patterns. In both, lack of information and cognitive bias are the primary sources that tempt people to believe the message (Boucher et al. 2021, Pennycook et al. 2020b). Both vaccine hesitancy and misinformation spread faster as they support intuition bias (Salali and Uysal 2020). Creating an echo chamber environment and homogenous social networks was proven to be "conducive to the spread of misinformation" (Jost et al. 2018). Following this background, it is prudent to believe that vaccine hesitancy messages like misinformation are designed based on the strategy of persuading readers to believe it.

Several studies have investigated public perception and sentiment of vaccine hesitancy messages through social media in the context of vaccine hesitancy. Still, a few studies went beyond that and analyzed and compared the strategies used in vaccine hesitancy messages with other vaccine messages. To study vaccine hesitancy messages and compare them with other COVID-19 messages, we need first to identify different messages, including hesitancy groups.

Therefore, this study aims to investigate three research questions. First, what are various heterogeneous topical groups that exist among COVID-19 vaccine-related social media messages? To answer this question, we developed a semantic network and applied Louvain modularity to group COVID-19 vaccine-related posts shared in the public Facebook groups. To label the groups and find the topics shared in each group, we created another network within each group. This network is based on bigrams co-occurrence of each group and helps find the most critical bigram in each group. The second research question is, do these communities exhibit certain network phenomena? To answer this question, we investigated the network characteristics of each group. The third question is, how are these groups different in terms of emotion?

The remainder of this contribution is provided as follows. Section 4.2 presents previous works investigating the posts about the COVID-19 vaccine on social media. In section 4.3, we describe the research methods used in this contribution. Section 4.4 and 4.5 provide the results and discussion, respectively. Finally, we provide the conclusion in section 4.6.

4.2 Related work

Several studies have investigated the attitudes toward vaccines and the reasons behind vaccine hesitancy. In general, based on the method of data collecting, these studies can be classified into two broad categories: traditional methods (such as questionnaires, surveys, and interviews) and infoveillance methods (such as social media). One of the common focuses of the traditional method was developing and applying a scale that can measure vaccine hesitancy and confidence. For example, Betsch et al. (2018) validated using the 5C scale in measuring the psychological antecedents of vaccination. In another example, De Figueiredo et al. (2020) assessed public perceptions of vaccine effectiveness, safety, and significance. Similarly, Batty et al. (2021) used individual-level data from 11,740 individuals to study the relationship between cognitive function and vaccine hesitancy.

Collecting data through traditional methods can be costly and time-consuming, while infoveillance sources can contain a wealth of real-time and historical data (Chew and Eysenbach 2010b). Advances in NLP methods make the infoveillance approach even more popular for both quantitative and qualitative analysis. As discussed, social media is one source of online information and infoveillance approach. Social media has been previously used in monitoring other pandemics. For example, Chew and Eysenbach (2010b) used social media data and applied content analysis to investigate public perceptions during the 2009 H1N1 Outbreak. In the current Covid-19 pandemic, many authors employ social media to analyze different aspects of vaccine acceptance and hesitancy attitudes. Sentiment and topical analysis are the two main parts of most infoveillance studies.

Lyu et al. (2022) studied public opinions on potential COVID-19 vaccines through six million tweets. They classified the tweets by a combination of manual coding and machine learning methods. They found that the opinions on the COVID-19 vaccine are different based on the personal characteristics of people, such as income, demographics, religion, and family status. Hou et al. (2021) monitored vaccine confidence, hesitancy, and public engagement in five global metropolises through Twitter and Sina Weibo. The posts were manually coded and classified into categories such as accept, neutral, doubt, and refuse. They found that negative posts are more about a lack of confidence and attract higher social engagement. Huangfu et al. (2022) applied sentiment analysis and topic modeling on a large dataset of tweets. They found the main concerns in the negative tweets are about vaccine hesitancy, side effects, and supply rollout. In another study focusing on sentiment analysis, Zhang et al. (2021) applied a sentiment analysis method and classified the vaccine-related posts on Weibo into positive and negative categories. They found that the rate of positive posts to negative posts fluctuated between 45% to 77% during the time window. Moreover, their findings showed that males are more positive than females about COVID-19 vaccines. Similarly, Chen et al. (2022) have applied clustering and LDA on a large dataset containing tweets to identify major topics.

Social network analysis is widely used in opinion mining. For example, Boucher et al. (2021) applied social network analysis, unsupervised learning, and BERT to find the latent topics in Twitter conversations. They identified four groups: vaccine acceptant, vaccine hesitant, Indian vaccine acceptant, and French vaccine hesitant groups. Similarly, Malova (2021) applied network and text analysis methods to Twitter data. They identified four main groups of stakeholders that governed Twitter discussions. Muric et al. (2021) collected tweets that show a robust anti-vaccine stance and estimated their political leaning. In addition, they applied topic analysis on anti-vaccine narratives by creating a hashtag co-occurrence network where each node is a hashtag and edges show the co-occurrence of hashtags. They found three focuses, including conspiracy, safety, and a mixture of various hashtags. Luo et al. (2021) investigated public perception toward the COVID-19 vaccine in two different countries: the United States and China. By creating a co-occurrence network analysis, they found the positive feeling about the vaccine in Weibo-based data from China and an anti-vaccine view on Twitter from the USA.

Most of the previous studies used social network analysis to study public perception and indentify the themes in the social media messages. In contrast, we implement a semantic network where the nodes are the posts and edges are semantic similarities. The advantage of this network is that the semantically similar posts can be identified and classified in the same group (topical groups). Moreover, to the best of our knowledge, few studies examine the network characteristics and emotions of different topical groups. These characteristics can reveal the strategies that anti-vaxers used in creating and disseminating their posts.

4.3 Research Method

In this study, we leveraged NLP and semantic network analysis methods to analyze Facebook posts shared on public groups related to the COVID-19 vaccine. The proposed method in this study is illustrated in Figure 4.1 and consists of three tiers. Tier 1 is our highest level of analysis, where we want to find the answer to the first research question. This tier includes data collection, semantic similarities analysis, and the creation of the semantic network. Tier 2 is our second level of analysis and is at the network level. At this level, we applied some network analysis methods including calculating clustering coefficient, average of shortest path and bridging centrality to measure the network characteristics and answer the second research question. Tier 3 is the word level of analysis, where we label each group and measure the emotion within the groups. Figure 4.1 shows the workflow of this study.

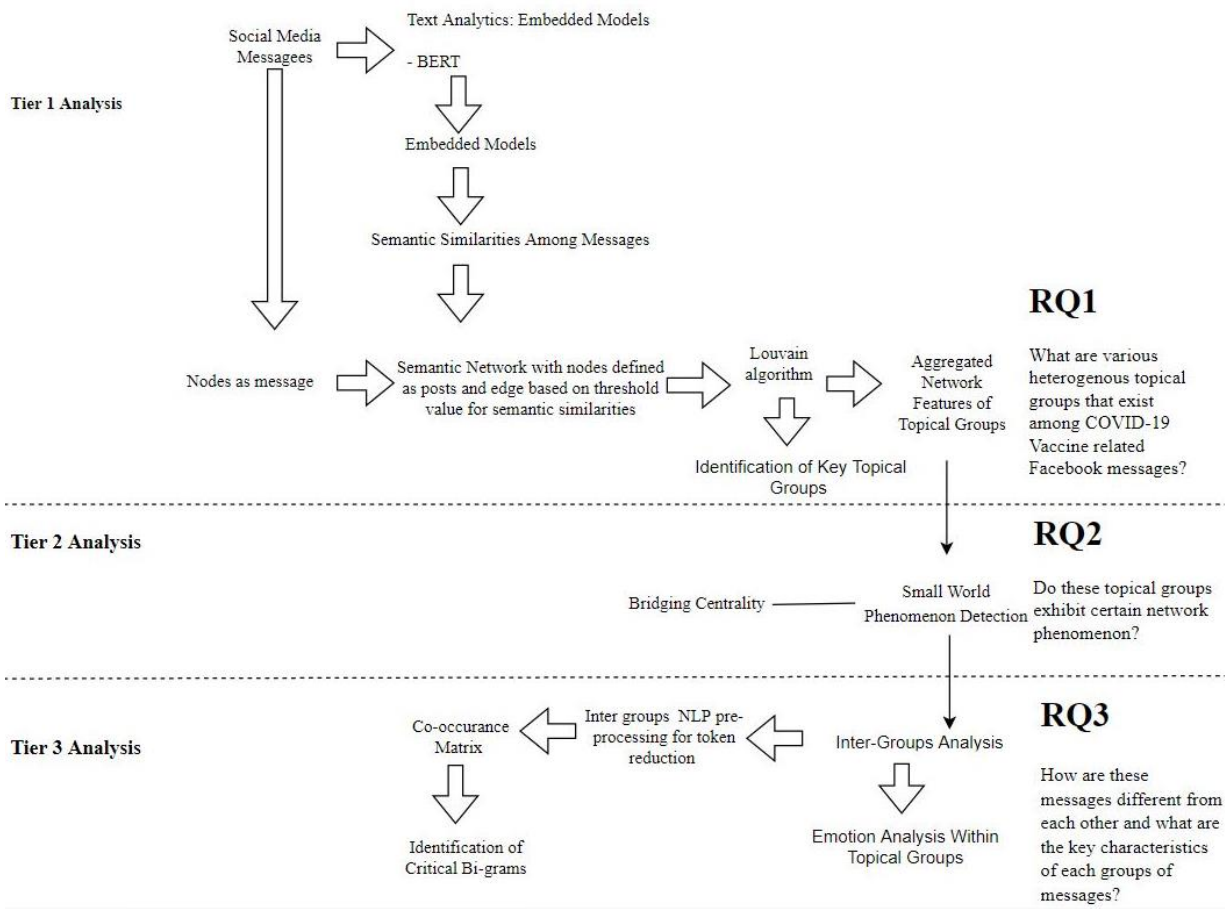

Figure 4.1: Workflow

4.3.1 Data collection and preparation

We draw our analysis based on a dataset obtained from the CrowdTangle web-based platform. CrowdTangle is a web tool that tracks posts published on public Facebook pages, groups, and verified profiles and provides the number of interactions for each post while excluding users from the dataset. We limited our study to a three-month window from January 2021 to March 2021 in the U.S. We filtered the inquiries to provide the English-language posts published in public Facebook groups where members shared their opinions and interests. We used the query (COVID, corona) AND (vaccine)) and collected 7480 posts from 2476 Facebook groups.

CrowdTangle provides the text from a Facebook post and provides the information in different categories, including message, link text, image text, and description. Before applying semantic similarities to Facebook posts, we combined the message, link text, and image text while excluding the description to avoid repetition. Figure 4.2 depicts a Facebook post to provide a better insight into how CrowdTangle provides the information from a Facebook post (The *i* sign is to show more information about the link).

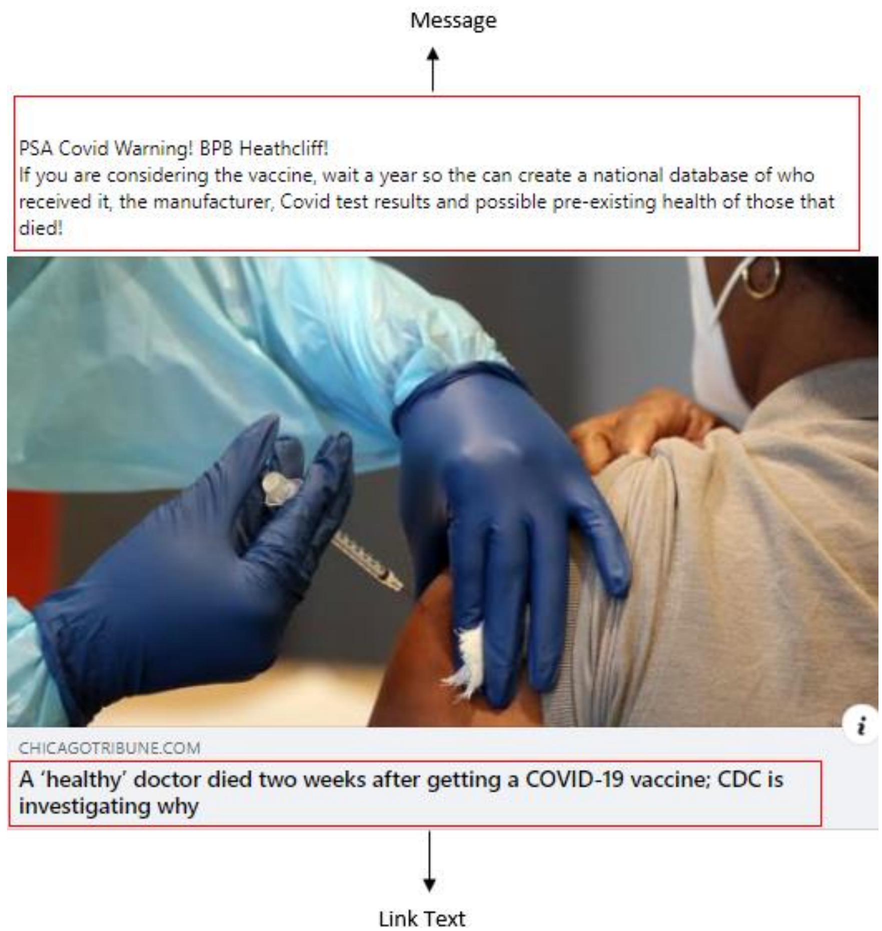

Figure 4.2: Facebook post

4.3.2 *Semantic Network Analysis and classifying the posts*

To characterize Facebook posts related to the COVID-19 vaccine, we created a semantic network where each node represents a post, and the edges indicate the semantic relation between the posts. Two nodes are connected if they have semantic similarities of more than 75%, and the percentages of similarities indicate the weight of the edge. We used trial-and-error tests to determine the threshold. We created a semantic network for each test and checked the modularity score. A higher modularity score means more distinguishable groups. The modularity score for a threshold of 72% was 0.81, while we achieved the modularity score of 0.66 and 0.57 for 70% and 65% threshold, respectively.

To find semantic similarities between posts, we leveraged cosine similarity methods to find the distance between the embedding space of posts. Embedding space was obtained through BERT. The simplest version of BERT (Devlin et al. 2018) is a pre-trained unsupervised language model that learns contextual relationships between the words on a large corpus. In this model, a Transformer, that is bidirectionally trained, encodes all words into a dense vector in the embedding space and bidirectionally. A Transformer (Vaswani et al. 2017) is an encoder-decoder architecture based on a neural network. The advantage of BERT is that, unlike the directional model, the Transformer encoder reads the entire text together and captures the context of each word based on the position of the given word and its surrounding words (Roy 2021). Therefore, the words that appear in a similar context will have a similar score. Other than providing embedding space, BERT can be applied to different NLP tasks (Chopard et al. 2021). In this study, we used Sentence-BERT (S-BERT) (Reimers and Gurevych 2019), which is a variant of BERT for generating sentence embeddings. This model leverage Siamese BERT networks (Chicco 2021) to use for tasks such as semantic similarity prediction. Given our training objectives, we employed “all-mpnet-base-v2” (Song et al. 2020) which is an S-BERT model pre-trained on a large-scale dataset (over 1B sentence pairs from multiple datasets), that can convert a text into a high-dimensional embedding vector. We leveraged the Transformer package of HuggingFace in Python to generate embeddings through this model.

Once we created the semantic network, we applied Louvain algorithm methods through Gephi software (Bastian et al. 2009) to classify the posts into different groups. Gephi is software that can be used for network visualization and analysis. The modularity function in Gephi uses the Louvain algorithm, which is an unsupervised clustering method. This function aims to find the optimum number of groups that maximizes the modularity score (Blondel et al. 2008). The modularity score measures the density of edges inside the groups with respect to edges between the groups.

4.3.3 Content Analysis

In addition to investigating each group manually to label the groups, we also applied some topic extraction methods to identify the topics that describe each group. There are different methods for extracting the topics and themes from the textual data. One of the methods is LDA which has attracted much attention from researchers. However, the problem with LDA is the number of topics that should be decided in advance, and often generated topics are uninterpretable (Dou et al. 2021). Recently network-based approaches have been used as methods of topic discovery in a text (Dou et al. 2021, You et al. 2016).

4.3.4 Co-occurrence network

Before constructing the co-occurrence networks, we applied some pre-processing steps to our textual data. The pre-processing included tokenization, stemming, removing stop words, converting all words to lower case, replacing some words with their synonym (like adverse event and side effect), and converting a word like side effect with side_effect. The co-occurrence networks were constructed where each node represents a bigram, and the weights of the edges indicate the frequencies of co-occurrence between two bigrams. A bigram is the combination of two adjacent words (like New York) and is a better unit of analysis in topic modeling (Wallach 2006).

After creating co-occurrence networks, we identified the topics shared in each group and labeled the groups. Each topic can be described by a set of keywords (Dou et al. 2021). These keywords can be identified by recognizing the centrality pattern of the words. The centrality pattern can indicate the significance and the role of a word in the text (Abbe and Falissard 2017, Yano et al. 2018). The most popular centrality measurements are degree, closeness, eigenvector, and betweenness centralities (You et al. 2016). Out of these, we used the weighted degree centrality, which is an extension of degree centrality.

4.3.5 Emotion Analysis

Previous studies showed the importance of emotion analysis by studying the association between emotion and vaccine decision or vaccine information flow (Christy et al. 2016, Himelboim et al. 2020). In the next step, we analyzed the emotion used in the language within each group. We adopted Ekman's six emotions as well as a neutral state. We employed the DistilRoBERTa base model (Hartmann 2022) to calculate Ekman's emotions and the neutral category. This model was trained on 6 different datasets.

4.4 Result

In this section, we describe the result of this study. First, we present the identified groups and then for each groups we show the co-occurrence network and key bi-gram. Then we provide the results of emotions and network analysis for each group.

4.4.1 Semantic network and clustering results

After constructing the semantic network on Gephi and running modularity, we found five giant clusters. Out of 19715 nodes, 7468 of them are included in our giant networks. Figure 4.3 shows the network, where these five clusters are shown in orange, blue, red, green, and pink. We analyzed the posts shared in each group and labeled them as mix, information, side effect, political and anti-vaccine. Table 4.1 indicates the number of nodes and edges in each cluster, and Table 4.2 describes the definition of the label in each group.

Groups	**Nodes**	**Edges**
Orange (Mix)	1923	23092
Blue (Information)	1858	24797
Green (Side effect)	1619	35800
Violet (Political)	1230	35564
Red (Anti-vaccine)	838	13927

Table 4.1: Size of each cluster

Label	**Description**
Information	Providing information about the vaccine Claiming the vaccine is effective Encouraging people to take the vaccine
Side effect	Having concerns regarding vaccine side effects Sharing experience about the vaccine side effects Not discouraging against the vaccine
Political	Talking about the role of government in vaccine development and distribution
Anti-vaccine	Providing information that misleads and discourages people about vaccine Claiming vaccine results in death and severe damage
Mix	Does not include a specific group

Table 4.2: Describing each label category

Figure 4.3: Network of five giant groups: Orange(mix), blue (information), green (side effects), violet (political) and red (anti-vaccine)

4.4.2 Co-occurrence network (Topic Identification) results

In this section, we provide the results of the co-occurrence network associated with the five groups (orange, blue, green, violet, and red).

4.4.2.1 Mix

Figure 4.4 represents a co-occurrence network built on the word from the largest group (orange group) and filtered by the top 15 highest weighted degree centrality words. The size of each node indicates a higher weighted degree. The biggest node shows the bigram with the highest connection with other nodes. Investigating this group in the semantic network along with the critical words in the co-occurrence leads us to the conclusion that this cluster does not include a specific group and is a mix of different groups. We concluded that although most of the posts in this cluster refer to anti-vaccine and vaccine-hesitancy views, there are posts that are pro-vaccine. Figure 4.5 depicts an example of posts in this group.

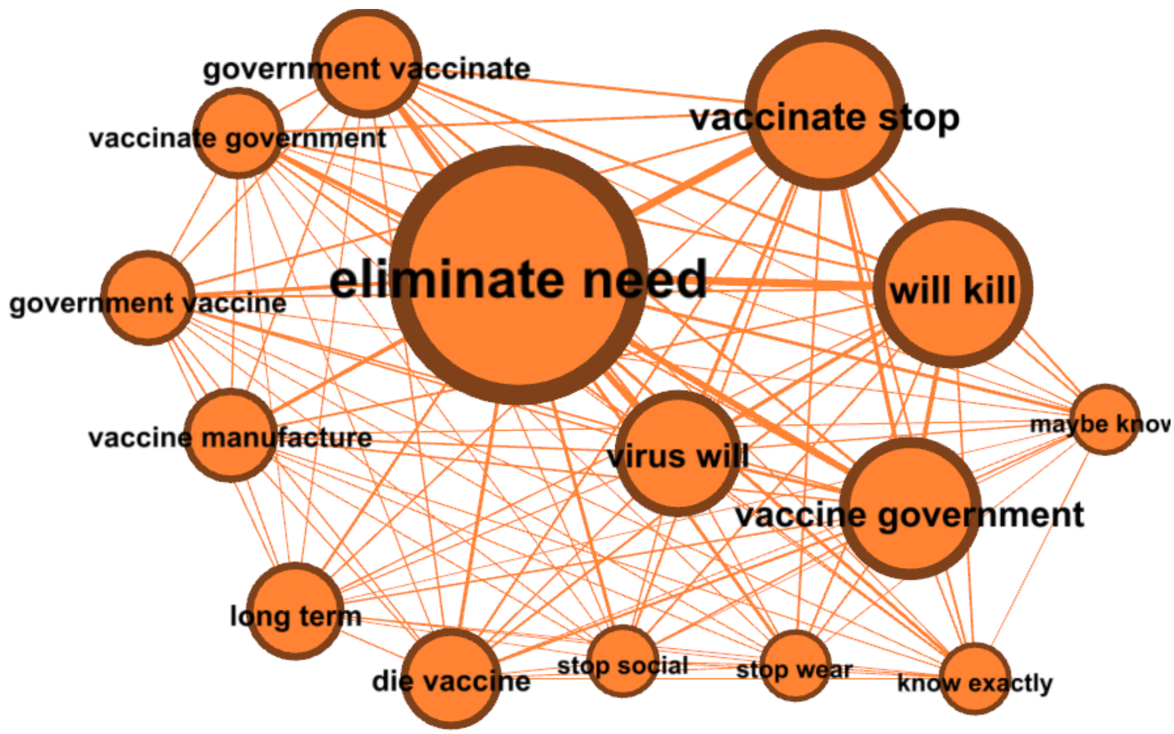

Figure 4.4: Top 15 Weighted degree for mix group

Figure 4.5: A post shared in the mix group

4.4.2.2 Information

Figure 4.6 shows the co-occurrence network of the second largest group (blue group) filtered by the highest weighted degree centrality (top 15 words). Investigating the co-occurrence network for this group indicates that most of the posts in this group provide general information about the vaccine. These posts emphasize vaccine efficiency, such as the post shown in Figure 4.7.

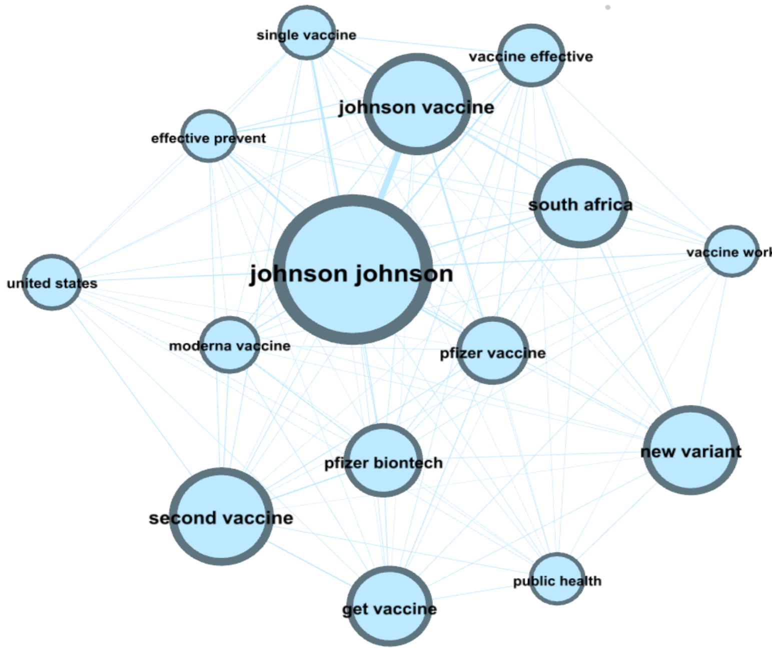

Figure 4.6: Top 15 Weighted degree for Information group

Johnson &Johnson's data is out, and, while the results don't match Pfizer's and Moderna's, the vaccine's one-shot advantage may go always in compensating, given the press now to get as many people vaccinated as possible. The company is expected to apply to the FDA for Emergency Authorization next week, and the vaccine could be authorized and available to the public late February-early March.
Here's the key points:
The vaccine was 85 percent effective at preventing severe disease, and 72 percent effective at protecting against moderate to severe illness in the United States, but it was 66 percent effective in Latin America and 57 percent effective in South Africa, where concerning variants have taken root. It was 85 percent effective overall at preventing severe disease.

In the clinical trial, all cases of covid-related hospitalization and death — the outcomes that most people would like to avoid — were among participants who had received placebo shots.

The logistical advantages of the Johnson & Johnson vaccine should not be underestimated, experts said, particularly at a time when the pursuit of broad immunity to the virus through vaccines has assumed new urgency with the emergence of variants, sparking concerns that mutant viruses might escape vaccines, treatments and naturally acquired immunity.

"Vaccinating fast is going to be the message now, and vaccinating fast to reduce transmission and reduce the chance you have additional mutations," said Paul Stoffels, chief scientific officer for Johnson & Johnson. "The crisis is now hospitalization and death."

WASHINGTONPOST.COM

Single-shot Johnson & Johnson vaccine 66 percent effective against moderate and severe illness

Figure 4.7: A post shared in the information group

4.4.2.3 Side effect

Figure 8 shows the co-occurrence network of the third largest group (green) filtered by the highest weighted degree centrality (top 15 words). This cluster includes the posts shared by people who are going to take the vaccine or have already taken it, and they are either concerned about side effects or shared the side effect(s) that they have experienced. The group include the post such as "Scheduled to get my first COVID Vaccine tomorrow. Curious if anyone has gotten theirs and if so, what the side-Effects were. On the fence about getting it, not sure...." or "How many have had vaccine that had the COVID virus? I had virus in July but getting my first injection today. Also I have antibodies. Did you have any reactions?".

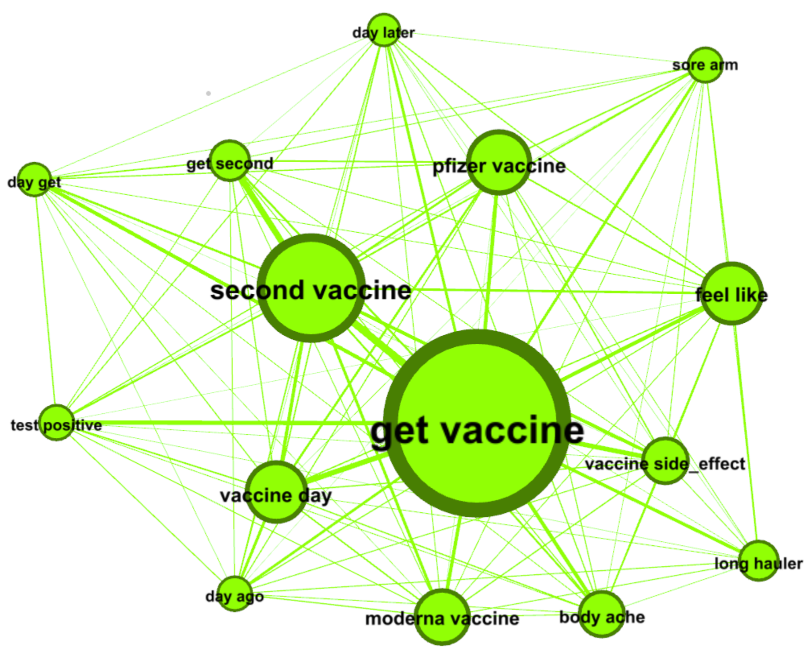

Figure 4.8: Top 15 Weighted degree for side effect group

4.4.2.4 Political

Figures 4.9 represent the co-occurrence network from the fourth group (violet) filtered by the top 15 bigrams in weighted degree centrality. Figure 4.10 depicts an example of this group. As the figures show, the bigrams in these networks are mostly related to political issues and the way the vaccine has been distributed by presidents.

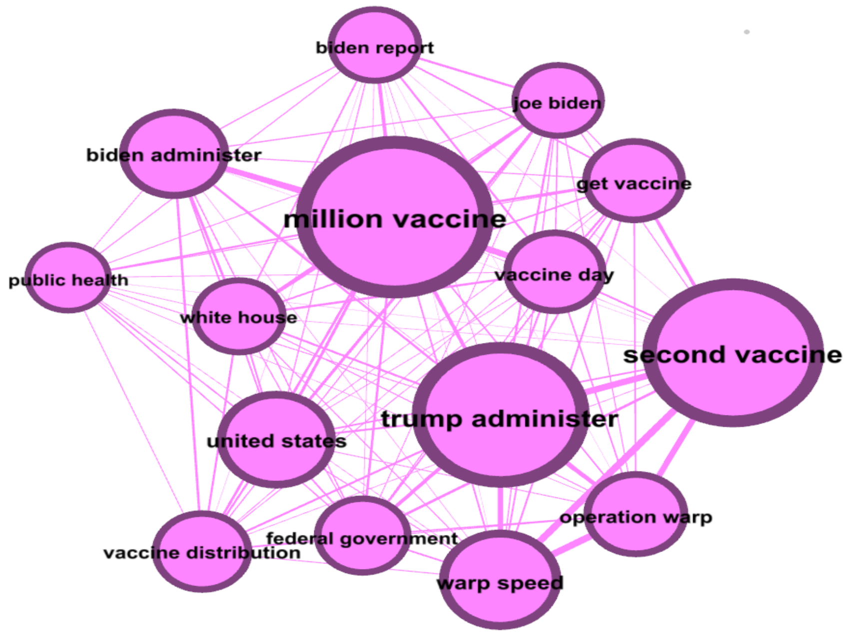

Figure 4.9: Top 15 Weighted degree for political group

- "The administration says it plans to buy an additional 100 million doses each from both #Moderna and #Pfizer, which has a vaccine with it's German counterpart, BioNTech.

"This increases the total vaccine order for the U.S. by 50%, from 400 million to 600 million with these additional doses expected to deliver this summer," the White House said in a fact sheet it put out before planned remarks from the president."

- "The Department of Health and Human Services will provide allocation estimates for the upcoming three weeks as opposed to the one week look-ahead that they previously received," the White House said in its fact sheet.

This increased transparency will give state and local leaders greater certainty around supply so that they can plan their vaccination efforts and administer vaccines effectively and efficiently."

- "Over the next three weeks, the Biden administration says it will boost supply to states, tribes and territories from a current 8.6 million doses a week to a minimum of 10 million doses per week."

- "The allocation and delivery of vaccines — now crossing two administrations — has been troubled. Some 20 million more doses have been distributed by the federal government than have actually been administered to people, according to the latest tally from the Centers for Disease Control and Prevention."

NPR.ORG

The Biden Administration Is Working To Buy 200 Million More COVID-19 Vaccine Doses

Figure 4.10 : A post shared in the political group

4.4.2.5 Anti-vaccine

Figure 4.11 represents the co-occurrence network from the last group (red) filtered by the top 15 weighted degree centrality. It can be seen that this group mainly includes words that refer to people who died after getting the vaccine, and the word vaccine had the most co-occurrence with the words such as die and side effects. The side effect in this group refers to serious side effects. Figure 4.12 depicts a post from this group.

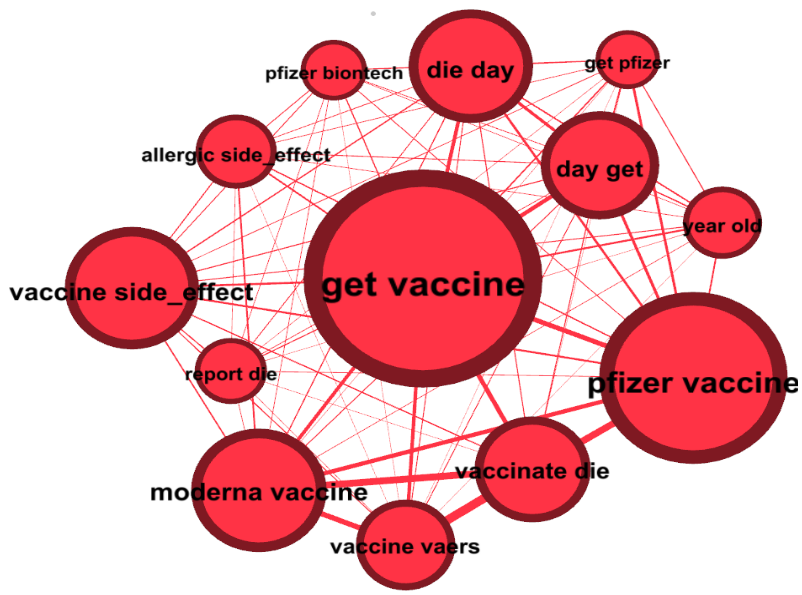

Figure 4.11: Top 15 Weighted degree for political group

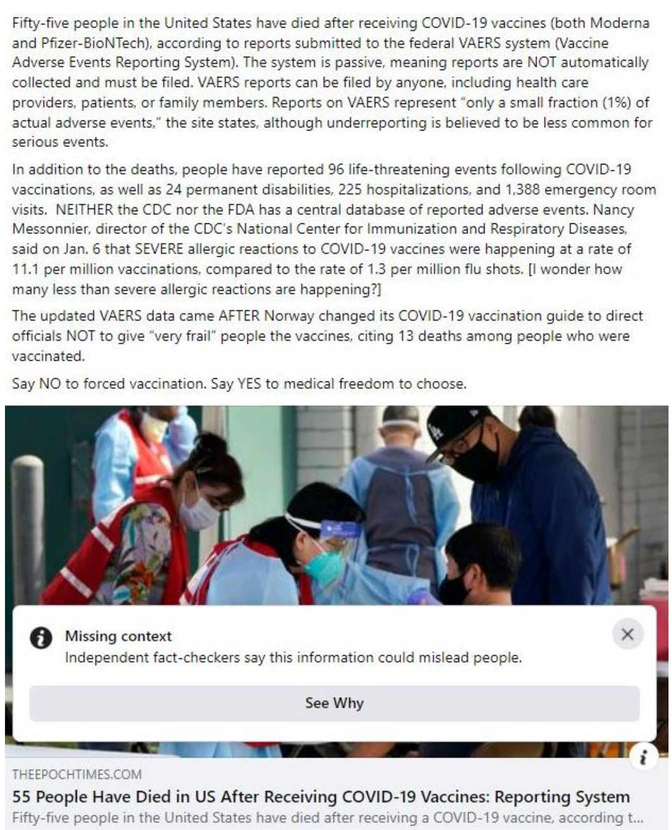

Figure 4.12: A post shared in the political group

4.4.3 *Emotion Analysis results*

We compared and contrasted different emotions within each group and between the groups. Table 4.3 displays the distribution of emotions in each group. Figures 4.13 and 4.14 visualize the distribution of emotions in each group and provide a comparison between the groups, respectively. We have excluded the mix group from further analysis as it does not provide information about a specific topic, though we show the information of this groups.

	Anger	Fear	Disgust	Sadness	Joy	Surprise	Neutral
Mix	0.115	0.228	0.040	0.070	0.053	0.092	0.402
Information	0.054	0.216	0.023	0.084	0.101	0.100	0.421
Side effect	0.023	0.187	0.038	0.121	0.136	0.113	0.383
Political	0.097	0.147	0.037	0.093	0.075	0.144	0.407
Anti-vaccine	0.041	0.237	0.020	0.375	0.016	0.110	0.201

Table 4.3: Distribution of emotions in the groups

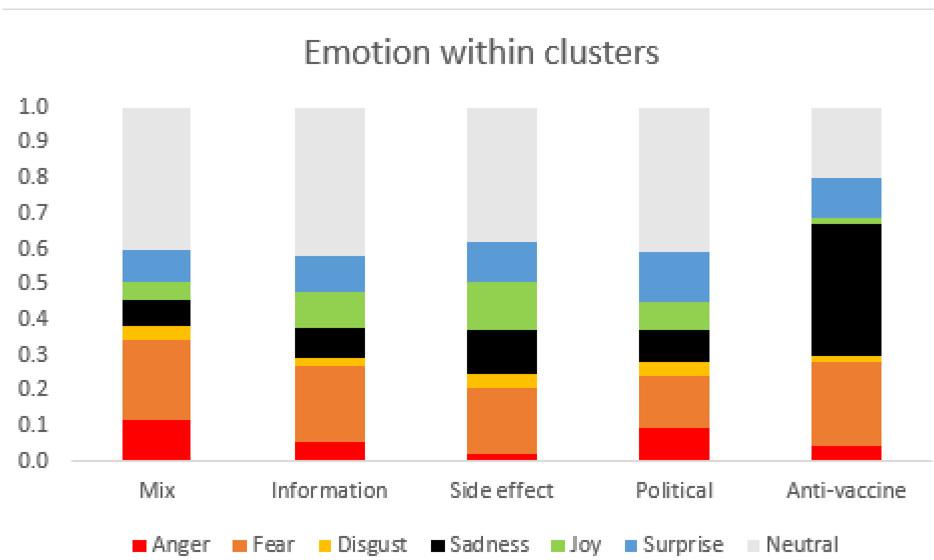

Figure 4.13: Distribution of emotion within groups

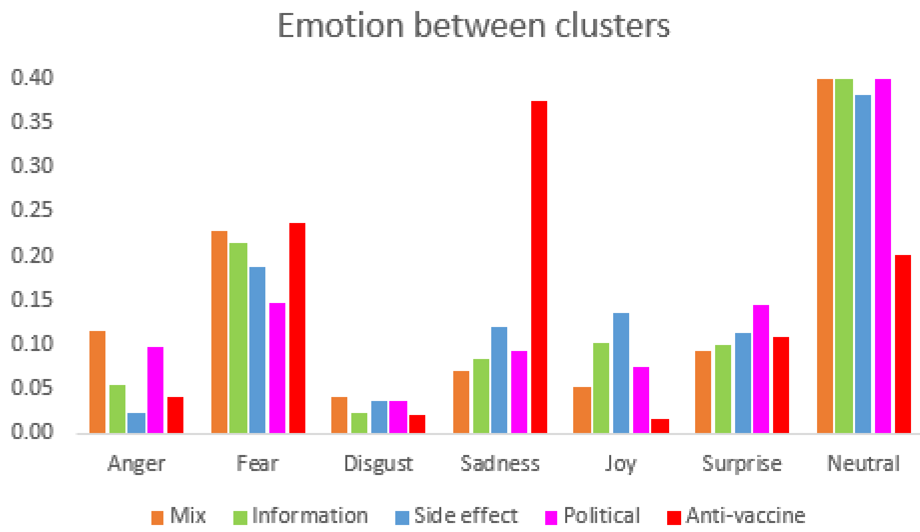

Figure 4.14: Emotion comparison between groups

4.4.4 Network characteristics results

In this section, we describe the network characteristics within each group and answer the question of whether these groups exhibit certain network phenomena. Network characteristics of these groups can provide important insights into how the posts in each group are created and linked together. First, we analyzed which group represents more characteristics of the small-world phenomenon (Telesford et al. 2011). A small-world network is a type of network with the characteristics of a high cluster coefficient and a small average of shortest path length. The clustering coefficient is the ratio of the number of edges between the neighbors of a given node and the maximum number of edges among those neighbors (Bosma et al. 2009). The average of the shortest path also indicates the shortest path between two nodes. In the context of a semantic network, a small-world network shows higher semantic similarities between all nodes in the group. Table 4.4 represents these two characteristics for each group.

Groups	Clustering Coefficient	Avg of the shortest path
Mix	0.615	3.744
Information	0.591	3.799
Side effect	0.531	3.177
Anti-vaccine	0.651	3.168
Political	0.654	2.801

Table 4.4: Clustering coefficient and an average of the shortest path within each group

Second, we calculated the bridging coefficient within each cluster and compared them. Bridging centrality is calculated by multiplying betweenness centrality by the bridging coefficient. The bridging coefficient of a node shows how well a given node is located between high-degree nodes (Macker 2016). In our case, a node with high bridging centrality is a node that links the parts of the network together. This node can connect different themes; for example, this node can be a post about the lack of proof that connects autism spectrum disorder to the vaccine (Bail 2016). Therefore, a group with a higher average of bridging centrality indicates more nodes with these properties. Table 4.5 shows the average bridging centrality within each cluster.

Cluster	Bridging Centrality
Mix	0.00004
Information	0.00005
Side effect	0.00004
Political	0.00003
Anti-vaccine	0.00001

Table 4.5: Bridging Centrality of each group

4.5 Discussion

We studied more than 7000 posts about the COVID vaccine shared on public groups on Facebook between January 2021 and March 2021. We raised two research objectives in this study. First, we wanted to check if there are heterogeneous topical groups of posts on Facebook. Second, we aimed to examine the different characteristics of these groups in terms of emotion and network.

The findings can provide insights for researchers and pro-vaccine organizations to design a strategy that might be more effective in combating anti-vaccine groups. In this section, we begin by describing the key findings and then we provide the practical implications, limitations, and future direction of this study.

4.5.1 Key Findings

We discovered 5 giant networks, which constitute 38% of the whole dataset. Out of these five groups, one group was a mix of different topics, and the others show more homogenous characteristics, meaning that the posts shared within these groups refer to a specific topic. One of these groups included the posts reflecting anti-vaccine views and attempts. The critical bigram in these groups revealed that the main themes are deadly damage and dying after receiving the vaccine. We found two homogenous groups that could be categorized as pro-vaccine groups. These groups were named information and side effects. The main themes around the information are about the efficacy of the vaccine and the need for taking the vaccine; also, the themes inside the side effect group are about possible side effects. The last group included the posts related to political issues, where the posts are in support of or against one of the presidents and their performance in vaccine development and distribution.

Next, we investigated the emotion used in the posts within each group. We also supported the emotion analysis results with a One-way analysis of variance (ANOVA) and Tukey's test. The results of Tukey's test are shown in section 6 (Appendix). Based on our findings, anti-vaccine posts were found to be more emotional than other groups, where the neutral state in the anti-vaccine group is significantly less than in other groups. Anti-vaccine posts also showed significantly less joy and more sadness than other groups. Fear in anti-vaccine and information groups were similar and significantly higher than in others. The political group was found to have higher angriness and surprise than other groups. Both pro-vaccine groups had joyful emotions that were significantly higher than others, in which the side effects groups had the highest amount of joy. The joy in the side effect group proved that the side effect concern in this group was not discouraging toward a vaccine. Among all emotions, disgust was the least dominant emotion.

Our findings using network analysis shows similar attributes between the political and anti-vaccine groups. Both these groups show more characteristics of a small-world network. In comparison with other groups, these two groups have a smaller average of the shortest path and a higher clustering coefficient. The smaller average shortest path shows that, the shorter semantic distance between the nodes and higher clustering coefficient shows higher semantic similarities between the nodes. Therefore, in the political group, only a few semantically different posts are connected. The small-world features in the political group is in line with the political polarization that we found in our theme analysis. Political polarization is not surprising in social media, and it has already been proven in previous research such as Gruzd et al. (2014). The anti-vaccine group is the second top group in terms of small-world features.

In addition, both the political and anti-vaccine groups have a lower bridging coefficient than the other groups, which supports the idea that a lower number of themes are involved in these groups and that these groups show more topical homophily, whereas the higher bridging coefficient in pro-vaccine groups indicates some nodes play a bridge role in combining different themes and cover a broader range of concern or information.

Therefore, topical homophily in the anti-vaccine group shows that the writers of anti-vaccine posts are targeting a specific audience. Adding the sadness and fear emotion to these posts shows that they targeted people who are skeptical or afraid about the vaccine. This finding also supports the idea that pro-vaccine groups target general readers and try to encourage the public to get a vaccine.

This behavior is very similar to COVID-19 fake news writers; Gupta et al. (2022) found that fake news writers target their audience and craft the message based on what they want.

4.5.2 Implications for research and practice

In this study, we found that anti-vaccine and political groups have similar network characteristics where both groups have more topical homophily than other groups. Future study can focus on these two groups to find out the proportion of misinformation that published in these groups.

In addition, this study helps pro-vaccine agencies identify the main themes in anti-vaccine posts and design interventions. Although other studies consider conspiracy theories as one of the main themes in anti-vaccine groups, in this study, we argue that characterizing vaccines as life-threatening is the main focus in the anti-vaccine groups. In addition, as we discussed above, the network characteristics of the anti-vaccine group are like fake news in the context of COVID-19 where both groups focus on narrow topics and craft the message for specific readers. Therefore, following the suggestions of Gupta et al. (2022) for a real news writer to write different messages for different audiences instead of writing a general message for the general audience, the pro-vaccine agencies should follow the same suggestion. For example, the information groups, as the name implies, provide general information about the vaccine, and inside the general information, they mention the need for quick vaccination. If pro-vaccine agencies focused on anti-vaccine groups, we should have found a group of posts talking specifically against anti-vaccines and saying that vaccines do not have serious side effects or justify the side effects.

4.6 Conclusion

Vaccine refusal is a serious threat in the recent global COVID-19 pandemic. Vaccination decisions can be affected by online information, and social media is counted as one of the main sources of online information. Therefore, it is important to monitor and study the information disseminated on social media. While a majority focus of the previous studies was themes analysis, in this study, we compared and contrasted the strategies that were used in different topical groups in addition to themes analysis of the information shared on Facebook.

To classify different information on Facebook, we applied a Louvain modularity on a semantic network where nodes and edges are Facebook posts and the semantic similarities between posts (with a threshold of 75%), respectively. Subsequently, we constructed a bigram words co-occurrence network and found the critical bigram within each group that represents the topic of each group.

We identified four heterogeneous groups of posts on Facebook between January 2021 till the end of March 2021. These groups are named Information, Side effect, Political, and Anti-vaccine. The first two groups belong to the pro-vaccine view, the political group is mostly about connecting vaccine development and distribution to the president, and anti-vaccine groups try to mislead people by introducing the COVID vaccine as a life-threatening vaccine.

Further, we examined and compared the emotion and network characteristics of the groups. Our findings suggest there are more small-world features in political and anti-vaccine groups than in others. These results suggest the existence of higher semantic similarities between all posts in the political and anti-vaccine groups. We also found that the average of bridging centrality is lower in anti-vaccine and political groups than in others, which shows the existence of a lower number of nodes that play a bridge role to connect different nodes. Therefore, we can conclude that there is more topical homophily in these two groups. These findings proves that the writers of posts in these groups focus on a specific audience. In addition, the results of emotion analysis also support that there are more emotional posts in the anti-vaccine group where fear and sadness are significantly higher while joy is significantly less than in other groups.

5 Conclusion

In the era of digital news, social media has become the primary source of information where people can easily access a vast amount of information. In addition, the information can quickly be disseminated and shared a million times through social media. Despite of the bright sides that these characteristics bring for social media, the dark sides cannot be ignored. One of the primary concerns regarding social media is that social media can be misused to shape public opinion.

In this dissertation, we implemented NLP and network analysis methods to study two types of information that are used in social media to shape people's opinions: fake news and anti-vaccine. Both of these types of information have always been in social media. However, they become more important after special events or crises. For example, fake news attracted attention in the 2016 election and increased again during COVID-19. Anti-vaccine information also starts rising during COVID-19.

The contribution of this dissertation is three-fold. In the first contribution, we analyzed general fake and satirical news and provided a recommendation system for the social media platforms. The proposed recommendation system can detect the type of news in the first stage, where we applied topic modeling methods such as LDA and LSA to extract the themes and topics from the news and implemented a random forest model to detect the type of news. In the second stage, the recommendation system can recommend a user the most similar real news based on the extracted topics (distribution of topics) and sentiment score of the latest read news by the user. Therefore, the proposed model, on the one hand, detects fake and satirical news. On the other hand, it recommends similar real news to avoid creating filter bubbles and echo chambers and reduce the speed of dissemination. The results of this study suggested that detecting the type of news based on the topics can achieve 85% accuracy. In addition, we evaluate the model with the other validation metrics such as sensitivity, specificity, precision, and f1-score and got a performance of over 80%.

This study suffers from some limitations. First, this study suffers from a lack of data that enables us to evaluate the results of our recommendation system. Therefore, future studies can find the data that help assess the recommendation system results. Second, the number of features considered for the recommendation system can be improved. For example, future studies can improve the recommendation model by adding more data and features such as location, age, sex, and implicit or explicit ratings of the user (which indicate the user's preferences).

In the second contribution, we focused on fake news related to COVID-19. This contribution is a theory-based study where we based our research on various deception theories. Particularly, we raised some hypotheses to investigate the different strategies the fake news writers might leverage in their writing in the context of COVID-19. We applied NLP methods to measure and PLS-SEM technique to test the hypotheses. We support the results by implementing MANOVA. The results of this contribution show that fake news in the context of COVID-19 has significantly more uncertainty, less complexity in their sentence, more complexity in vocabulary level, less specificity, more negative affect, less diversity, less information quantity, more expressivity of words, and more cognitive process. We evaluated the discovered hypotheses by creating a detection model and applying the XGBoost model to the above results as the model's features. The detection model achieved 70% accuracy.

This study suffers from several limitations. The first limitation of this study refers to the dataset that we collected. To label the data, we had to rely on other studies, such as (Gupta et al., 2022), since manual labeling in the big dataset is very hard and requires experts who can distinguish fake news from real news on the subject of COVID-19. The next limitation refers to the number of articles that we were able to collect. We found a small number of fake news articles to collect. We only had 4476 fake news articles after cleaning, which is not a large dataset on the scale of big data, and due to the limitation with fake news, we collected the same number of real news to keep our dataset balanced. Another limitation of this study is the types of news we were able to collect. While some other types of false news, such as satire, can be categorized as false news, we limited our studies to fake news. Future studies can develop this research from three perspectives: dataset, theory, and methods. Regarding dataset, as mentioned, we did not cover different categories of false news. Therefore, it could be interesting to test if the claimed hypothesis is true for other types of false news (like satire) or not. Moreover, in the future, we can categorize the dataset in terms of topics such as politics, health care, and opinion to test if the subject of articles can affect the result of the hypothesis or not. Regarding theory, future studies can develop linguistics features and psychological theory. This study mostly focused on communication deception theories, while future studies can cover broader theories that bring other hypotheses to test. Regarding the methods, future studies can develop the method by applying a more sophisticated model such as variation autoencoder as the predictive model where variational autoencoder can reveal important features.

In the last contribution, we focused on anti-vaccine messages shared on Facebook. In this contribution, we aimed to answer three research objectives. First: we sought to find the various heterogenous topical groups in Facebook posts related to COVID-19. We expected to find anti-vaccine groups along with some other groups. To this end, all posts related to COVID-19 vaccine that were shared on Facebook public groups between January 2021 and the end of March 2021 were collected and analyzed semantically using a BERT model. Then a semantic network was created to classify the semantically similar posts. The result of this part suggests 5 big topical groups. Subsequently, we created a bigram co-occurrence network within each group and filtered the top 15 words with the highest weighted degree to find the themes of the groups. The second research objective is to find the network characteristics and phenomena in the anti-vaccine groups and compare them with other groups. This result suggests that the anti-vaccine and political groups have more attributes of a small-world network and less bridging centrality, proving that these groups focused on more homogenous topics and targeted a specific audience. Also, anti-vaccine groups are significantly more emotional, less joyful, and less sad than other groups.

This contribution suffered from some limitations. First, we excluded around 60% of our data as they were not included in our giant groups. Second, we only collected three months' data due to computing power limitations. Third, our data did not include the comments in each post, where the comments could provide great insight into the readers' reaction to each post. Fourth, our data did not include any user information while investigating the user who shares anti-vaccine posts could provide a better analysis.

In the future, the anti-vaccine posts' text characteristics (word diversity, text complexity) can be investigated and compared with other groups. In addition, machine learning and network analysis methods can be integrated to support semantic similarities and add more data to the giant groups.

6 Appendix

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
Anti_vaccine	Information	0.0124	0.1891	-0.0032	0.028	False
Anti_vaccine	Mix	0.0739	0.0	0.0585	0.0894	True
Anti_vaccine	Political	0.0556	0.0	0.0389	0.0723	True
Anti_vaccine	Side effect	-0.0183	0.0146	-0.0342	-0.0024	True
Information	Mix	0.0615	0.0	0.0494	0.0737	True
Information	Political	0.0432	0.0	0.0295	0.0569	True
Information	Side effect	-0.0307	0.0	-0.0434	-0.018	True
Mix	Political	-0.0183	0.0023	-0.032	-0.0047	True
Mix	Side effect	-0.0922	0.0	-0.1048	-0.0796	True
Political	Side effect	-0.0739	0.0	-0.088	-0.0598	True

Figure 6.1: Tukey's test of anger emotion

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
Anti_vaccine	Information	-0.021	0.4044	-0.0538	0.0118	False
Anti_vaccine	Mix	-0.0085	0.9543	-0.0411	0.0241	False
Anti_vaccine	Political	-0.0901	0.0	-0.1254	-0.0548	True
Anti_vaccine	Side effect	-0.0499	0.0005	-0.0834	-0.0164	True
Information	Mix	0.0125	0.6704	-0.0131	0.0381	False
Information	Political	-0.0691	0.0	-0.0981	-0.0402	True
Information	Side effect	-0.0289	0.0267	-0.0557	-0.0021	True
Mix	Political	-0.0816	0.0	-0.1104	-0.0529	True
Mix	Side effect	-0.0414	0.0002	-0.068	-0.0149	True
Political	Side effect	0.0402	0.0022	0.0104	0.07	True

Figure 6.2: The results of tukey's test for fear emotion

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
Anti_vaccine	Information	0.0035	0.8658	-0.0063	0.0133	False
Anti_vaccine	Mix	0.0204	0.0	0.0106	0.0301	True
Anti_vaccine	Political	0.0175	0.0001	0.007	0.028	True
Anti_vaccine	Side effect	0.0178	0.0	0.0078	0.0278	True
Information	Mix	0.0169	0.0	0.0092	0.0245	True
Information	Political	0.014	0.0001	0.0053	0.0226	True
Information	Side effect	0.0143	0.0	0.0063	0.0223	True
Mix	Political	-0.0029	0.8905	-0.0115	0.0057	False
Mix	Side effect	-0.0026	0.9029	-0.0105	0.0054	False
Political	Side effect	0.0003	1.0	-0.0086	0.0092	False

Figure 6.3: The results of tukey's test for disgust emotion

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
Anti_vaccine	Information	-0.2905	0.0	-0.3125	-0.2684	True
Anti_vaccine	Mix	-0.3048	0.0	-0.3268	-0.2829	True
Anti_vaccine	Political	-0.2822	0.0	-0.3059	-0.2584	True
Anti_vaccine	Side effect	-0.2538	0.0	-0.2763	-0.2312	True
Information	Mix	-0.0144	0.1535	-0.0316	0.0029	False
Information	Political	0.0083	0.7718	-0.0112	0.0278	False
Information	Side effect	0.0367	0.0	0.0187	0.0547	True
Mix	Political	0.0227	0.0121	0.0033	0.042	True
Mix	Side effect	0.0511	0.0	0.0332	0.069	True
Political	Side effect	0.0284	0.0011	0.0083	0.0484	True

Figure 6.4: The results of tukey's test for sadness emotion

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
Anti_vaccine	Information	0.0847	0.0	0.0638	0.1056	True
Anti_vaccine	Mix	0.0364	0.0	0.0156	0.0572	True
Anti_vaccine	Political	0.0588	0.0	0.0363	0.0813	True
Anti_vaccine	Side effect	0.1195	0.0	0.0981	0.1409	True
Information	Mix	-0.0483	0.0	-0.0646	-0.0319	True
Information	Political	-0.0259	0.0012	-0.0444	-0.0074	True
Information	Side effect	0.0348	0.0	0.0177	0.0519	True
Mix	Political	0.0224	0.0079	0.004	0.0407	True
Mix	Side effect	0.0831	0.0	0.0661	0.1	True
Political	Side effect	0.0607	0.0	0.0417	0.0797	True

Figure 6.5: The results of tukey's test for joy emotion

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
Anti_vaccine	Information	-0.0091	0.6694	-0.0278	0.0095	False
Anti_vaccine	Mix	-0.0175	0.076	-0.0361	0.0011	False
Anti_vaccine	Political	0.0346	0.0	0.0144	0.0547	True
Anti_vaccine	Side effect	0.0032	0.9911	-0.0159	0.0223	False
Information	Mix	-0.0084	0.5213	-0.023	0.0062	False
Information	Political	0.0437	0.0	0.0272	0.0602	True
Information	Side effect	0.0123	0.178	-0.0029	0.0276	False
Mix	Political	0.0521	0.0	0.0357	0.0685	True
Mix	Side effect	0.0207	0.0018	0.0056	0.0359	True
Political	Side effect	-0.0314	0.0	-0.0484	-0.0144	True

Figure 6.6: The results of tukey's test for suprise emotion

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
Anti_vaccine	Information	0.22	0.0	0.185	0.255	True
Anti_vaccine	Mix	0.2001	0.0	0.1653	0.2349	True
Anti_vaccine	Political	0.2058	0.0	0.1681	0.2435	True
Anti_vaccine	Side effect	0.1815	0.0	0.1457	0.2173	True
Information	Mix	-0.0199	0.2734	-0.0473	0.0075	False
Information	Political	-0.0142	0.7204	-0.0451	0.0167	False
Information	Side effect	-0.0385	0.0022	-0.0671	-0.0099	True
Mix	Political	0.0057	0.9866	-0.025	0.0364	False
Mix	Side effect	-0.0186	0.3795	-0.047	0.0098	False
Political	Side effect	-0.0243	0.2261	-0.0562	0.0075	False

Figure 6.7: The results of tukey's test for neutral emotion

7 References

Abbe, A. & Falissard, B. 2017. 'Stopping antidepressants and anxiolytics as major concerns reported in online health communities: a text mining approach.' *JMIR mental health,* 4:4, e7797.

Addawood, A., Badawy, A., Lerman, K. & Ferrara, E. 2019. 'Linguistic cues to deception: Identifying political trolls on social media.' Paper presented at Proceedings of the International AAAI Conference on Web and Social Media.

AGNUS, R. 2020. 'Institute. More Canadians willing to roll up their sleeves right away as national COVID-19 vaccine rollout begins-Angus Reid Institute [Internet]. 2020.'

Ahmed, W., Vidal-Alaball, J., Downing, J. & Seguí, F. L. 2020. 'COVID-19 and the 5G conspiracy theory: social network analysis of Twitter data.' *Journal of Medical Internet Research,* 22:5, e19458.

Al-Rakhami, M. S. & Al-Amri, A. M. 2020. 'Lies Kill, Facts Save: Detecting COVID-19 Misinformation in Twitter.' *IEEE Access,* 8, 155961-70.

Allcott, H. & Gentzkow, M. 2017. 'Social media and fake news in the 2016 election.' *Journal of economic perspectives,* 31:2, 211-36.

Amatriain, X., Jaimes, A., Oliver, N. & Pujol, J. M. 2011. 'Data mining methods for recommender systems.' *Recommender systems handbook:* 39-71. Springer.

Antoun, W., Baly, F., Achour, R., Hussein, A. & Hajj, H. 2020. 'State of the Art Models for Fake News Detection Tasks.' Paper presented at 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT).

Apuke, O. D. & Omar, B. 2020. 'User motivation in fake news sharing during the COVID-19 pandemic: an application of the uses and gratification theory.'

Ashraf, M., Tahir, G. A., Abrar, S., Abdulaali, M., Mushtaq, S. & Mukthar, H. 2018. 'Personalized news recommendation based on multi-agent framework using social media preferences.' Paper presented at 2018 International Conference on Smart Computing and Electronic Enterprise (ICSCEE).

Aslam, S. 2020. 'Twitter by the Numbers: Stats, Demographics & Fun Facts.'

Bail, C. A. 2016. 'Combining natural language processing and network analysis to examine how advocacy organizations stimulate conversation on social media.' *Proceedings of the National Academy of Sciences,* 113:42, 11823-28.

Barthel, M., Mitchell, A. & Holcomb, J. 2016. 'Many Americans believe fake news is sowing confusion.' *Pew research center,* 15, 12.

Basch, C. H. & MacLean, S. A. 2019. 'A content analysis of HPV related posts on instagram.' *Human vaccines & immunotherapeutics,* 15:7-8, 1476-78.

Bastian, M., Heymann, S. & Jacomy, M. 2009. 'Gephi: an open source software for exploring and manipulating networks.' Paper presented at Proceedings of the international AAAI conference on web and social media.

Batty, G. D., Deary, I. J., Fawns-Ritchie, C., Gale, C. R., Altschul, D. J. B., behavior, & immunity 2021. 'Pre-pandemic cognitive function and COVID-19 vaccine hesitancy: cohort study.' 96, 100-05.

Bessi, A. & Ferrara, E. 2016. 'Social bots distort the 2016 US Presidential election online discussion.' *First Monday,* 21:11-7.

Betsch, C., Renkewitz, F., Betsch, T. & Ulshöfer, C. 2010. 'The influence of vaccine-critical websites on perceiving vaccination risks.' *Journal of health psychology,* 15:3, 446-55.

Betsch, C., Schmid, P., Heinemeier, D., Korn, L., Holtmann, C. & Böhm, R. J. P. o. 2018. 'Beyond confidence: Development of a measure assessing the 5C psychological antecedents of vaccination.' 13:12, e0208601.

Bharadwaj, P. & Shao, Z. 2019. 'Fake news detection with semantic features and text mining.' *International Journal on Natural Language Computing (IJNLC) Vol,* 8.

Blei, D. M., Ng, A. Y. & Jordan, M. I. 2003. 'Latent dirichlet allocation.' *Journal of machine Learning research,* 3:Jan, 993-1022.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. 2008. 'Fast unfolding of communities in large networks.' *Journal of statistical mechanics: theory and experiment,* 2008:10, P10008.

Bodoff, D. & Ho, S. Y. 2016. 'Partial least squares structural equation modeling approach for analyzing a model with a binary indicator as an endogenous variable.' *Communications of the Association for Information Systems,* 38:1, 23.

Bosma, I., Reijneveld, J. C., Klein, M., Douw, L., Van Dijk, B. W., Heimans, J. J. & Stam, C. J. 2009. 'Disturbed functional brain networks and neurocognitive function in low-grade glioma patients: a graph theoretical analysis of resting-state MEG.' *Nonlinear biomedical physics,* 3:1, 1-11.

Boucher, J.-C., Cornelson, K., Benham, J. L., Fullerton, M. M., Tang, T., Constantinescu, C., Mourali, M., Oxoby, R. J., Marshall, D. A. & Hemmati, H. J. J. i. 2021. 'Analyzing social media to explore the attitudes and behaviors following the announcement of successful COVID-19 vaccine trials: infodemiology study.' 1:1, e28800.

Brady, W. J., Gantman, A. P. & Van Bavel, J. J. 2020. 'Attentional capture helps explain why moral and emotional content go viral.' *Journal of Experimental Psychology: General,* 149:4, 746.

Breiman, L. 2001. 'Random forests.' *Machine learning,* 45:1, 5-32.

Buller, D. B. & Burgoon, J. K. 1996. 'Interpersonal deception theory.' *Communication theory,* 6:3, 203-42.

Burgoon, J. K., Blair, J. P., Qin, T. & Nunamaker, J. F. 2003. 'Detecting deception through linguistic analysis.' Paper presented at International Conference on Intelligence and Security Informatics.

Burkhardt, J. M. 2017. 'History of fake news.' *Library Technology Reports,* 53:8, 5-9.

Busari;, S. & Adebayo, B. 2020. 'Nigeria records chloroquine poisoning after Trump endorses it for coronavirus treatment.'

Campos, P. G., Díez, F. & Cantador, I. 2014. 'Time-aware recommender systems: a comprehensive survey and analysis of existing evaluation protocols.' *User Modeling and User-Adapted Interaction,* 24:1-2, 67-119.

Caserotti, M., Girardi, P., Rubaltelli, E., Tasso, A., Lotto, L., Gavaruzzi, T. J. S. s. & medicine 2021. 'Associations of COVID-19 risk perception with vaccine hesitancy over time for Italian residents.' 272, 113688.

Chen, E., Jiang, J., Chang, H.-C. H., Muric, G. & Ferrara, E. 2022. 'Charting the information and misinformation landscape to characterize misinfodemics on social media: COVID-19 infodemiology study at a planetary scale.' *Jmir Infodemiology,* 2:1, e32378.

Chen, H.-C. & Chen, A. L. 2001. 'A music recommendation system based on music data grouping and user interests.' Paper presented at Proceedings of the tenth international conference on Information and knowledge management.

Chew, C. & Eysenbach, G. 2010a. 'Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak.' *PLoS One,* 5:11, e14118.

Chew, C. & Eysenbach, G. J. P. o. 2010b. 'Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak.' 5:11, e14118.

Chicco, D. J. A. N. N. 2021. 'Siamese neural networks: An overview.' 73-94.

Chopard, D., Treder, M. S., Corcoran, P., Ahmed, N., Johnson, C., Busse, M. & Spasic, I. J. J. M. I. 2021. 'Text mining of adverse events in clinical trials: Deep learning approach.' 9:12, e28632.

Christy, S. M., Winger, J. G., Raffanello, E. W., Halpern, L. F., Danoff-Burg, S. & Mosher, C. E. 2016. 'The role of anticipated regret and health beliefs in HPV vaccination intentions among young adults.' *Journal of Behavioral Medicine,* 39:3, 429-40.

Ciampaglia, G. L., Shiralkar, P., Rocha, L. M., Bollen, J., Menczer, F. & Flammini, A. 2015. 'Computational fact checking from knowledge networks.' *PLoS One,* 10:6, e0128193.

Cleger-Tamayo, S., Fernández-Luna, J. M. & Huete, J. F. 2012. 'Top-N news recommendations in digital newspapers.' *Knowledge-based systems,* 27, 180-89.

Conroy, N. K., Rubin, V. L. & Chen, Y. 2015. 'Automatic deception detection: Methods for finding fake news.' *Proceedings of the Association for Information Science and Technology,* 52:1, 1-4.

Cover, T. M. & Hart, P. 1967. 'Nearest neighbor pattern classification.' *IEEE transactions on information theory,* 13:1, 21-27.

Daley, B. P. 2020. 'Leveraging Machine Learning for Automatically Classifying Fake News in the COVID-19 Outbreak.'

De Figueiredo, A., Simas, C., Karafillakis, E., Paterson, P. & Larson, H. J. J. T. L. 2020. 'Mapping global trends in vaccine confidence and investigating barriers to vaccine uptake: a large-scale retrospective temporal modelling study.' 396:10255, 898-908.

De Sarkar, S., Yang, F. & Mukherjee, A. 2018. 'Attending sentences to detect satirical fake news.' Paper presented at Proceedings of the 27th International Conference on Computational Linguistics.

del Pilar Salas-Zárate, M., Paredes-Valverde, M. A., Rodriguez-García, M. Á., Valencia-García, R. & Alor-Hernández, G. 2017. 'Automatic detection of satire in Twitter: A psycholinguistic-based approach.' *Knowledge-based systems,* 128, 20-33.

Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. J. a. p. a. 2018. 'Bert: Pre-training of deep bidirectional transformers for language understanding.'

Dhama, K., Sharun, K., Tiwari, R., Dhawan, M., Emran, T. B., Rabaan, A. A., Alhumaid, S. J. H. V. & Immunotherapeutics 2021. 'COVID-19 vaccine hesitancy–reasons and solutions to achieve a successful global vaccination campaign to tackle the ongoing pandemic.' 17:10, 3495-99.

Dharawat, A., Lourentzou, I., Morales, A. & Zhai, C. 2020. 'Drink bleach or do what now? Covid-HeRA: A dataset for risk-informed health decision making in the presence of COVID19 misinformation.' *arXiv preprint arXiv:2010.08743*.

Diamantopoulos, A. & Winklhofer, H. M. 2001a. 'Index construction with formative indicators: An alternative to scale development.' *Journal of marketing research,* 38:2, 269-77.

Diamantopoulos, A. & Winklhofer, H. M. J. J. o. m. r. 2001b. 'Index construction with formative indicators: An alternative to scale development.' 38:2, 269-77.

Dou, M., Wang, Y., Gu, Y., Dong, S., Qiao, M., Deng, Y. J. C. & Geosciences 2021. 'Disaster damage assessment based on fine-grained topics in social media.' 156, 104893.

Feng, C., Khan, M., Rahman, A. U. & Ahmad, A. 2020. 'News Recommendation Systems-Accomplishments, Challenges & Future Directions.' *IEEE Access,* 8, 16702-25.

Feng, S., Banerjee, R. & Choi, Y. 2012. 'Syntactic stylometry for deception detection.' Paper presented at Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers).

Fine, P., Eames, K. & Heymann, D. L. J. C. i. d. 2011. '"Herd immunity": a rough guide.' 52:7, 911-16.

Fortuna, B., Fortuna, C. & Mladenić, D. 2010. 'Real-time news recommender system.' Paper presented at Joint European Conference on Machine Learning and Knowledge Discovery in Databases.

Fuller, C. M., Biros, D. P., Burgoon, J. & Nunamaker, J. 2013. 'An examination and validation of linguistic constructs for studying high-stakes deception.' *Group decision and negotiation,* 22:1, 117-34.

Geschke, D., Lorenz, J. & Holtz, P. 2019. 'The triple-filter bubble: Using agent-based modelling to test a meta-theoretical framework for the emergence of filter bubbles and echo chambers.' *British Journal of Social Psychology,* 58:1, 129-49.

Ghenai, A. & Mejova, Y. 2017. 'Catching Zika fever: Application of crowdsourcing and machine learning for tracking health misinformation on Twitter.' *arXiv preprint arXiv:1707.03778*.

Golbeck, J., Mauriello, M., Auxier, B., Bhanushali, K. H., Bonk, C., Bouzaghrane, M. A., Buntain, C., Chanduka, R., Cheakalos, P. & Everett, J. B. 2018. 'Fake news vs satire: A dataset and analysis.' Paper presented at Proceedings of the 10th ACM Conference on Web Science.

Goodman, B. 2022. 'Is herd immunity for Covid-19 still possible?'.

Gravelle, T. B., Phillips, J. B., Reifler, J., Scotto, T. J. J. H. V. & Immunotherapeutics 2022. 'Estimating the size of "anti-vax" and vaccine hesitant populations in the US, UK, and Canada: comparative latent class modeling of vaccine attitudes.' 1-10.

Greene, D., O'Callaghan, D. & Cunningham, P. 2014. 'How many topics? stability analysis for topic models.' Paper presented at Joint European Conference on Machine Learning and Knowledge Discovery in Databases.

Grice, P. 1989. *Studies in the Way of Words.* Harvard University Press.

Gruzd, A., Roy, J. J. P. & internet 2014. 'Investigating political polarization on Twitter: A Canadian perspective.' 6:1, 28-45.

Gupta, A., Li, H., Farnoush, A. & Jiang, W. 2022. 'Understanding patterns of COVID infodemic: A systematic and pragmatic approach to curb fake news.' *Journal of business research,* 140, 670-83.

Haim, M., Graefe, A. & Brosius, H.-B. 2018. 'Burst of the filter bubble? Effects of personalization on the diversity of Google News.' *Digital journalism,* 6:3, 330-43.

Hair Jr, J. F., Hult, G. T. M., Ringle, C. & Sarstedt, M. 2016. *A primer on partial least squares structural equation modeling (PLS-SEM).* Sage publications.

Hancock, J. T., Curry, L. E., Goorha, S. & Woodworth, M. J. D. P. 2007. 'On lying and being lied to: A linguistic analysis of deception in computer-mediated communication.' 45:1, 1-23.

Hardalov, M., Koychev, I. & Nakov, P. 2016. 'In search of credible news.' Paper presented at International Conference on Artificial Intelligence: Methodology, Systems, and Applications*.*

Hartmann, J. 2022. 'Emotion English DistilRoBERTa-base.'

Hauch, V., Blandón-Gitlin, I., Masip, J. & Sporer, S. L. 2015. 'Are computers effective lie detectors? A meta-analysis of linguistic cues to deception.' *Personality and social psychology Review,* 19:4, 307-42.

Hauver, D. B. & French, J. C. 2001. 'Flycasting: using collaborative filtering to generate a playlist for online radio.' Paper presented at Proceedings First International Conference on WEB Delivering of Music. WEDELMUSIC 2001*.*

Himelboim, I., Xiao, X., Lee, D. K. L., Wang, M. Y. & Borah, P. 2020. 'A social networks approach to understanding vaccine conversations on Twitter: Network clusters, sentiment, and certainty in HPV social networks.' *Health communication,* 35:5, 607-15.

Ho, S. M., Hancock, J. T., Booth, C. & Liu, X. 2016. 'Computer-mediated deception: Strategies revealed by language-action cues in spontaneous communication.' *Journal of Management Information Systems,* 33:2, 393-420.

Honnibal, M. 2017. 'Spacy.'

Honnibal, M. & Montani, I. 2017. 'spacy 2: Natural language understanding with bloom embeddings, convolutional neural networks and incremental parsing.' *To appear,* 7:1.

Horne, B. D. & Adali, S. 2017. 'This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news.' *arXiv preprint arXiv:1703.09398*.

Hossain, T., Logan IV, R. L., Ugarte, A., Matsubara, Y., Young, S. & Singh, S. 2020. 'COVIDLIES: Detecting COVID-19 Misinformation on Social Media.' Paper presented at Proceedings of the 1st Workshop on NLP for COVID-19 (Part 2) at EMNLP 2020*.*

Hosseinimotlagh, S. & Papalexakis, E. E. 2018. 'Unsupervised content-based identification of fake news articles with tensor decomposition ensembles.' Paper presented at Proceedings of the Workshop on Misinformation and Misbehavior Mining on the Web (MIS2)*.*

Hou, Z., Tong, Y., Du, F., Lu, L., Zhao, S., Yu, K., Piatek, S. J., Larson, H. J. & Lin, L. J. J. o. m. I. r. 2021. 'Assessing COVID-19 vaccine hesitancy, confidence, and public engagement: A global social listening study.' 23:6, e27632.

Huangfu, L., Mo, Y., Zhang, P., Zeng, D. D. & He, S. J. J. o. m. I. r. 2022. 'COVID-19 Vaccine Tweets After Vaccine Rollout: Sentiment–Based Topic Modeling.' 24:2, e31726.

Humpherys, S. L., Moffitt, K. C., Burns, M. B., Burgoon, J. K. & Felix, W. F. 2011. 'Identification of fraudulent financial statements using linguistic credibility analysis.' *Decision Support Systems,* 50:3, 585-94.

Hyland, K. 1998. *Hedging in scientific research articles.* John Benjamins Publishing.

Islam, A. N., Laato, S., Talukder, S. & Sutinen, E. 2020. 'Misinformation sharing and social media fatigue during COVID-19: An affordance and cognitive load perspective.' *Technological Forecasting and Social Change,* 159, 120201.

Jaccard, J. & Wan, C. K. 1995. 'Measurement error in the analysis of interaction effects between continuous predictors using multiple regression: Multiple indicator and structural equation approaches.' *Psychological bulletin,* 117:2, 348.

Jadhav, S. S. & Thepade, S. D. 2019. 'Fake news identification and classification using DSSM and improved recurrent neural network classifier.' *Applied Artificial Intelligence,* 33:12, 1058-68.

Jamieson, K. H. & Cappella, J. N. 2008. *Echo chamber: Rush Limbaugh and the conservative media establishment.* Oxford University Press.

Jawad, M. R., Qasim, M. A., Cazzato, G., Zahra, M. M. A., Kapula, P. R., Gherabi, N., Jaleel, R. A. J. P. o. E. & Sciences, N. 2021. 'Advancement of artificial intelligence techniques based lexicon emotion analysis for vaccine of COVID-19.' 9:4, 580-88.

Jin, Y., Liu, B. F. & Austin, L. L. 2014. 'Examining the role of social media in effective crisis management: The effects of crisis origin, information form, and source on publics' crisis responses.' *Communication research,* 41:1, 74-94.

Johnson, M. K. & Raye, C. L. 1981. 'Reality monitoring.' *Psychological review,* 88:1, 67.

Jomsri, P. 2014. 'Book recommendation system for digital library based on user profiles by using association rule.' Paper presented at Fourth edition of the International Conference on the Innovative Computing Technology (INTECH 2014).

Jost, J. T., van der Linden, S., Panagopoulos, C. & Hardin, C. D. J. C. o. i. p. 2018. 'Ideological asymmetries in conformity, desire for shared reality, and the spread of misinformation.' 23, 77-83.

Karafillakis, E., Martin, S., Simas, C., Olsson, K., Takacs, J., Dada, S., Larson, H. J. J. J. p. h. & surveillance 2021. 'Methods for social media monitoring related to vaccination: systematic scoping review.' 7:2, e17149.

Karimi, H., Roy, P., Saba-Sadiya, S. & Tang, J. 2018a. 'Multi-source multi-class fake news detection.' Paper presented at Proceedings of the 27th International Conference on Computational Linguistics.

Karimi, M., Jannach, D. & Jugovac, M. 2018b. 'News recommender systems–Survey and roads ahead.' *Information Processing & Management,* 54:6, 1203-27.

Kestenbaum, L. A. & Feemster, K. A. J. P. a. 2015. 'Identifying and addressing vaccine hesitancy.' 44:4, e71-e75.

Khanday, A. M. U. D., Khan, Q. R. & Rabani, S. T. 2020. 'Identifying propaganda from online social networks during COVID-19 using machine learning techniques.' *International Journal of Information Technology*, 1-8.

Kim, J., Bae, J. & Hastak, M. 2018. 'Emergency information diffusion on online social media during storm Cindy in US.' *International Journal of Information Management,* 40, 153-65.

Kouzy, R., Abi Jaoude, J., Kraitem, A., El Alam, M. B., Karam, B., Adib, E., Zarka, J., Traboulsi, C., Akl, E. W. & Baddour, K. 2020. 'Coronavirus goes viral: quantifying the COVID-19 misinformation epidemic on Twitter.' *Cureus,* 12:3.

Kuzelewska, U. 2014. 'Clustering algorithms in hybrid recommender system on movielens data.' *Studies in logic, grammar and rhetoric,* 37:1, 125-39.

Laato, S., Islam, A. N., Islam, M. N. & Whelan, E. 2020. 'What drives unverified information sharing and cyberchondria during the COVID-19 pandemic?' *European Journal of Information Systems*, 1-18.

Lahariya, C. J. J. o. F. M. & Care, P. 2016. 'Vaccine epidemiology: A review.' 5:1, 7.

Leask, J., Willaby, H. W., Kaufman, J. J. H. v. & immunotherapeutics 2014. 'The big picture in addressing vaccine hesitancy.' 10:9, 2600-02.

Lee, H. J. & Park, S. J. 2007. 'MONERS: A news recommender for the mobile web.' *Expert Systems with Applications,* 32:1, 143-50.

Lex, E., Wagner, M. & Kowald, D. 2018. 'Mitigating confirmation bias on twitter by recommending opposing views.' *arXiv preprint arXiv:1809.03901*.

Li, F. 2008. 'Annual report readability, current earnings, and earnings persistence.' *Journal of Accounting and economics,* 45:2-3, 221-47.

Lim, X. J., Radzol, A., Cheah, J. & Wong, M. W. J. A. J. o. B. R. 2017. 'The impact of social media influencers on purchase intention and the mediation effect of customer attitude.' 7:2, 19-36.

Lin, F., Zhou, X. & Zeng, W. 2016. 'Sparse online learning for collaborative filtering.' *International Journal of Computers Communications & Control,* 11:2, 248-58.

Linden, G., Smith, B. & York, J. 2003. 'Amazon. com recommendations: Item-to-item collaborative filtering.' *IEEE Internet computing,* 7:1, 76-80.

Liu, J., Dolan, P. & Pedersen, E. R. 2010. 'Personalized news recommendation based on click behavior.' Paper presented at Proceedings of the 15th international conference on Intelligent user interfaces.

Liu, X., Agarwal, S., Ding, C. & Yu, Q. 2016. 'An LDA-SVM active learning framework for web service classification.' Paper presented at 2016 IEEE International Conference on Web Services (ICWS).

Lopez, G. 2016. 'Pizzagate, the Fake News Conspiracy Theory that Led a Gunman to DC's Comet Ping Pong, Explained.' *Vox, http://www. vox. com/policy-and-politics/2016/12/5/13842258/pizzagate-comet-ping-pong-fake-news*.

Lovari, A. 2020. 'Spreading (dis) trust: Covid-19 misinformation and government intervention in Italy.' *Media and Communication,* 8:2, 458-61.

Lu, Z., Dou, Z., Lian, J., Xie, X. & Yang, Q. 2015. 'Content-based collaborative filtering for news topic recommendation.' Paper presented at AAAI.

Luo, C., Chen, A., Cui, B. & Liao, W. 2021. 'Exploring public perceptions of the COVID-19 vaccine online from a cultural perspective: Semantic network analysis of two social media platforms in the United States and China.' *Telematics and Informatics,* 65, 101712.

Lyu, H., Wang, J., Wu, W., Duong, V., Zhang, X., Dye, T. D. & Luo, J. J. I. m. 2022. 'Social media study of public opinions on potential COVID-19 vaccines: informing dissent, disparities, and dissemination.' 2:01, 1-12.

MacKenzie, S. B., Podsakoff, P. M. & Jarvis, C. B. 2005. 'The problem of measurement model misspecification in behavioral and organizational research and some recommended solutions.' *Journal of applied psychology,* 90:4, 710.

Macker, J. P. 2016. 'An improved local bridging centrality model for distributed network analytics.' Paper presented at MILCOM 2016-2016 IEEE Military Communications Conference.

Malova, E. J. C. R. R. 2021. 'Understanding online conversations about COVID-19 vaccine on Twitter: vaccine hesitancy amid the public health crisis.' 38:5, 346-56.

Mangal, A. & Holm, E. A. 2018. 'A comparative study of feature selection methods for stress hotspot classification in materials.' *Integrating Materials and Manufacturing Innovation,* 7:3, 87-95.

Marr, B. 2020. 'Coronavirus Fake News: How Facebook, Twitter, And Instagram Are Tackling The Problem.' *Forbes, Mar*.

Masip, J., Sporer, S. L., Garrido, E. & Herrero, C. 2005. 'The detection of deception with the reality monitoring approach: A review of the empirical evidence.' *Psychology, Crime & Law,* 11:1, 99-122.

Maxouris, C. & Vera, A. 2021. 'There's no need to delay second dose of Covid-19 vaccine so more can get the first, Fauci says.'

McCornack, S. A., Levine, T. R., Solowczuk, K. A., Torres, H. I. & Campbell, D. M. 1992. 'When the alteration of information is viewed as deception: An empirical test of information manipulation theory.' *Communications Monographs,* 59:1, 17-29.

McKee, C., Bohannon, K. J. T. j. o. p. p. & therapeutics 2016. 'Exploring the reasons behind parental refusal of vaccines.' 21:2, 104-09.

McQuillan, L., McAweeney, E., Bargar, A. & Ruch, A. 2020. 'Cultural Convergence: Insights into the behavior of misinformation networks on Twitter.' *arXiv preprint arXiv:2007.03443*.

Mian, A. & Khan, S. 2020. 'Coronavirus: the spread of misinformation.' *BMC medicine,* 18:1, 1-2.

Mitra, T., Wright, G. P. & Gilbert, E. 2017. 'A parsimonious language model of social media credibility across disparate events.' Paper presented at Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing.

Mohseni, S. & Ragan, E. 2018. 'Combating Fake News with Interpretable News Feed Algorithms.' *arXiv preprint arXiv:1811.12349*.

Mooney, R. J. & Roy, L. 2000. 'Content-based book recommending using learning for text categorization.' Paper presented at Proceedings of the fifth ACM conference on Digital libraries.

Moran, M. B., Lucas, M., Everhart, K., Morgan, A. & Prickett, E. 2016. 'What makes anti-vaccine websites persuasive? A content analysis of techniques used by anti-vaccine websites to engender anti-vaccine sentiment.' *Journal of Communication in Healthcare,* 9:3, 151-63.

Muric, G., Wu, Y. & Ferrara, E. 2021. 'COVID-19 Vaccine Hesitancy on Social Media: Building a Public Twitter Data Set of Antivaccine Content, Vaccine Misinformation, and Conspiracies.' *JMIR public health and surveillance,* 7:11, e30642.

Mutanga, M. B. & Abayomi, A. 2020. 'Tweeting on COVID-19 pandemic in South Africa: LDA-based topic modelling approach.' *African Journal of Science, Technology, Innovation and Development*, 1-10.

Natarajan, S. & Moh, M. 2016. 'Recommending news based on hybrid user profile, popularity, trends, and location.' Paper presented at 2016 international conference on collaboration technologies and systems (CTS).

Newman, M. L., Pennebaker, J. W., Berry, D. S. & Richards, J. M. 2003. 'Lying words: Predicting deception from linguistic styles.' *Personality and social psychology bulletin,* 29:5, 665-75.

Oberlader, V. A., Naefgen, C., Koppehele-Gossel, J., Quinten, L., Banse, R. & Schmidt, A. F. 2016. 'Validity of content-based techniques to distinguish true and fabricated statements: A meta-analysis.' *Law and human behavior,* 40:4, 440.

Ofcom 2020. 'Half of UK adults exposed to false claims about coronavirus.'

Organization, W. H. 2019. 'Ten threats to global health in 2019.' World Health Organization.

Organization, W. H. 2020. 'Coronavirus disease (COVID-19) advice for the public: myth busters.'

Oyeyemi, S. O., Gabarron, E. & Wynn, R. 2014. 'Ebola, Twitter, and misinformation: a dangerous combination?' *Bmj,* 349, g6178.

Pan, S. L. & Zhang, S. 2020. 'From fighting COVID-19 pandemic to tackling sustainable development goals: An opportunity for responsible information systems research.' *International Journal of Information Management*, 102196.

Pariser, E. 2011. *The filter bubble: What the Internet is hiding from you.* Penguin UK.

Parizi, A. H. & Kazemifard, M. 2015. 'Emotional news recommender system.' Paper presented at 2015 Sixth International Conference of Cognitive Science (ICCS).

Pavlinek, M. & Podgorelec, V. 2017. 'Text classification method based on self-training and LDA topic models.' *Expert Systems with Applications,* 80, 83-93.

Pazzani, M. J. & Billsus, D. 2007. 'Content-based recommendation systems.' *The adaptive web:* 325-41. Springer.

Peden, B. F. & Carroll, D. W. 2008. 'Ways of writing: Linguistic analysis of self-assessment and traditional assignments.' *Teaching of Psychology,* 35:4, 313-18.

Pennebaker, J. W., Boyd, R. L., Jordan, K. & Blackburn, K. 2015. 'The development and psychometric properties of LIWC2015.'

Pennycook, G., McPhetres, J., Zhang, Y., Lu, J. G. & Rand, D. G. 2020a. 'Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention.' *Psychological science,* 31:7, 770-80.

Pennycook, G., McPhetres, J., Zhang, Y., Lu, J. G. & Rand, D. G. J. P. s. 2020b. 'Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention.' 31:7, 770-80.

Phorasim, P. & Yu, L. 2017. 'Movies recommendation system using collaborative filtering and k-means.' *International Journal of Advanced Computer Research,* 7:29, 52.

Piltch-Loeb, R., Savoia, E., Goldberg, B., Hughes, B., Verhey, T., Kayyem, J., Miller-Idriss, C. & Testa, M. 2021. 'Examining the effect of information channel on COVID-19 vaccine acceptance.' *PLoS One,* 16:5, e0251095.

Qureshi, S. M. & Kang, C. 2015. 'Analysing the organizational factors of project complexity using structural equation modelling.' *International Journal of Project Management,* 33:1, 165-76.

Raamkumar, A. S., Tan, S. G. & Wee, H. L. J. J. o. m. I. r. 2020. 'Measuring the outreach efforts of public health authorities and the public response on Facebook during the COVID-19 pandemic in early 2020: cross-country comparison.' 22:5, e19334.

Ramos, J. 2003. 'Using tf-idf to determine word relevance in document queries.' Paper presented at Proceedings of the first instructional conference on machine learning.

Rashkin, H., Choi, E., Jang, J. Y., Volkova, S. & Choi, Y. 2017. 'Truth of varying shades: Analyzing language in fake news and political fact-checking.' Paper presented at Proceedings of the 2017 conference on empirical methods in natural language processing.

Recasens, M., Danescu-Niculescu-Mizil, C. & Jurafsky, D. 2013. 'Linguistic models for analyzing and detecting biased language.' Paper presented at Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers).

Reimers, N. & Gurevych, I. J. a. p. a. 2019. 'Sentence-bert: Sentence embeddings using siamese bert-networks.'

Reis, J. C., Correia, A., Murai, F., Veloso, A. & Benevenuto, F. 2019. 'Supervised learning for fake news detection.' *IEEE Intelligent Systems,* 34:2, 76-81.

Reiter, P. L., Gerend, M. A., Gilkey, M. B., Perkins, R. B., Saslow, D., Stokley, S., Tiro, J. A., Zimet, G. D. & Brewer, N. T. J. A. p. 2018. 'Advancing human papillomavirus vaccine delivery: 12 priority research gaps.' 18:2, S14-S16.

Ribeiro, M. T., Ziviani, N., Moura, E. S. D., Hata, I., Lacerda, A. & Veloso, A. 2014. 'Multiobjective pareto-efficient approaches for recommender systems.' *ACM Transactions on Intelligent Systems and Technology (TIST),* 5:4, 1-20.

Ringle, C., Wende, S. & Becker, J.-M. 2015. 'SmartPLS 3. SmartPLS GmbH, Boenningstedt.' *Journal of Service Science and Management,* 10:3.

Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M. & Rigol-Sanchez, J. P. 2012. 'An assessment of the effectiveness of a random forest classifier for land-cover classification.' *ISPRS Journal of Photogrammetry and Remote Sensing,* 67, 93-104.

Roozenbeek, J., Schneider, C. R., Dryhurst, S., Kerr, J., Freeman, A. L., Recchia, G., van der Bles, A. M. & van der Linden, S. 2020. 'Susceptibility to misinformation about COVID-19 around the world.' *Royal Society Open Science,* 7:10, 201199.

Roy, A. 2021. 'Combining Text Embedding with Additional Knowledge for Information Extraction.' University of Maryland, Baltimore County.

Rubin, V. L., Chen, Y. & Conroy, N. K. 2015. 'Deception detection for news: three types of fakes.' *Proceedings of the Association for Information Science and Technology,* 52:1, 1-4.

Rubin, V. L. & Conroy, N. 2012. 'Discerning truth from deception: Human judgments and automation efforts.' *First Monday,* 17:5.

Rubin, V. L., Conroy, N., Chen, Y. & Cornwell, S. 2016. 'Fake news or truth? using satirical cues to detect potentially misleading news.' Paper presented at Proceedings of the second workshop on computational approaches to deception detection.

Rubin, V. L., Liddy, E. D. & Kando, N. 2006. 'Certainty identification in texts: Categorization model and manual tagging results.' *Computing attitude and affect in text: Theory and applications:* 61-76. Springer.

Salali, G. D. & Uysal, M. S. J. P. m. 2020. 'COVID-19 vaccine hesitancy is associated with beliefs on the origin of the novel coronavirus in the UK and Turkey.' 1-3.

Sallam, M., Dababseh, D., Eid, H., Al-Mahzoum, K., Al-Haidar, A., Taim, D., Yaseen, A., Ababneh, N. A., Bakri, F. G. & Mahafzah, A. J. V. 2021. 'High rates of COVID-19 vaccine hesitancy and its association with conspiracy beliefs: a study in Jordan and Kuwait among other Arab countries.' 9:1, 42.

Samantray, A. & Pin, P. 2019. 'Credibility of climate change denial in social media.' *palgrave communications,* 5:1, 1-8.

Saquete, E., Tomás, D., Moreda, P., Martínez-Barco, P. & Palomar, M. 2020. 'Fighting post-truth using natural language processing: A review and open challenges.' *Expert Systems with Applications,* 141, 112943.

Savov, P., Jatowt, A. & Nielek, R. 2020. 'Identifying breakthrough scientific papers.' *Information Processing & Management,* 57:2, 102168.

Schafer, J. B., Frankowski, D., Herlocker, J. & Sen, S. 2007. 'Collaborative filtering recommender systems.' *The adaptive web:* 291-324. Springer.

Schein, A. I., Popescul, A., Ungar, L. H. & Pennock, D. M. 2002. 'Methods and metrics for cold-start recommendations.' Paper presented at Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval.

Schütze, H., Manning, C. D. & Raghavan, P. 2008. *Introduction to information retrieval.* Cambridge University Press Cambridge.

Sehl, K. 2019. '27 Top Facebook Demographics that Matter to Social Media Marketers.'

Serrano, J. C. M., Papakyriakopoulos, O. & Hegelich, S. 2020. 'NLP-based Feature Extraction for the Detection of COVID-19 Misinformation Videos on YouTube.' Paper presented at Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020.

Sha, H., Hasan, M. A., Mohler, G. & Brantingham, P. J. 2020. 'Dynamic topic modeling of the COVID-19 Twitter narrative among US governors and cabinet executives.' *arXiv preprint arXiv:2004.11692*.

Shahi, G. K. & Nandini, D. 2020. 'FakeCovid--A Multilingual Cross-domain Fact Check News Dataset for COVID-19.' *arXiv preprint arXiv:2006.11343*.

Sharma, K., Qian, F., Jiang, H., Ruchansky, N., Zhang, M. & Liu, Y. 2019. 'Combating fake news: A survey on identification and mitigation techniques.' *ACM Transactions on Intelligent Systems and Technology (TIST),* 10:3, 1-42.

Shearer, E. 2018. 'Social media outpaces print newspapers in the US as a news source.' *Pew research center,* 10.

Shi, B. & Weninger, T. 2016. 'Discriminative predicate path mining for fact checking in knowledge graphs.' *Knowledge-based systems,* 104, 123-33.

Shokeen, J. & Rana, C. 2020. 'A study on features of social recommender systems.' *Artificial Intelligence Review,* 53:2, 965-88.

Shu, K., Cui, L., Wang, S., Lee, D. & Liu, H. 2019. 'defend: Explainable fake news detection.' Paper presented at Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining.

Shu, K., Sliva, A., Wang, S., Tang, J. & Liu, H. 2017. 'Fake news detection on social media: A data mining perspective.' *ACM SIGKDD Explorations Newsletter,* 19:1, 22-36.

Siering, M., Koch, J.-A. & Deokar, A. V. 2016. 'Detecting fraudulent behavior on crowdfunding platforms: The role of linguistic and content-based cues in static and dynamic contexts.' *Journal of Management Information Systems,* 33:2, 421-55.

Sindermann, C., Cooper, A. & Montag, C. 2020. 'A short Review on Susceptibility to Falling for Fake Political News.' *Current Opinion in Psychology*.

Singh, L., Bansal, S., Bode, L., Budak, C., Chi, G., Kawintiranon, K., Padden, C., Vanarsdall, R., Vraga, E. & Wang, Y. 2020. 'A first look at COVID-19 information and misinformation sharing on Twitter.' *arXiv preprint arXiv:2003.13907*.

Song, K., Tan, X., Qin, T., Lu, J. & Liu, T.-Y. 2020. 'Mpnet: Masked and permuted pre-training for language understanding.' *Advances in Neural Information Processing Systems,* 33, 16857-67.

Spinney, L. 2017. 'How Facebook, fake news and friends are warping your memory.' *Nature News,* 543:7644, 168.
Strapparava, C. & Mihalcea, R. 2008. 'Learning to identify emotions in text.' Paper presented at Proceedings of the 2008 ACM symposium on Applied computing.
Tandoc Jr, E. C., Lim, Z. W. & Ling, R. 2018. 'Defining “fake news” A typology of scholarly definitions.' *Digital journalism,* 6:2, 137-53.
Tausczik, Y. R. & Pennebaker, J. W. 2010. 'The psychological meaning of words: LIWC and computerized text analysis methods.' *Journal of language and social psychology,* 29:1, 24-54.
Tavakolifard, M., Gulla, J. A., Almeroth, K. C., Ingvaldesn, J. E., Nygreen, G. & Berg, E. 2013. 'Tailored news in the palm of your hand: a multi-perspective transparent approach to news recommendation.' Paper presented at Proceedings of the 22nd International Conference on World Wide Web.
Telesford, Q. K., Joyce, K. E., Hayasaka, S., Burdette, J. H. & Laurienti, P. J. 2011. 'The ubiquity of small-world networks.' *Brain connectivity,* 1:5, 367-75.
Toma, C. L. & Hancock, J. T. 2010. 'Reading between the lines: linguistic cues to deception in online dating profiles.' Paper presented at Proceedings of the 2010 ACM conference on Computer supported cooperative work.
Trujillo, K. L. & Motta, M. J. T. C. 2020. 'A majority of vaccine skeptics plan to refuse a COVID-19 vaccine, a study suggests, and that could be a big problem.' 10-06.
Ullman, J. B. & Bentler, P. M. 2003. 'Structural equation modeling.' *Handbook of psychology*, 607-34.
Undeutsch, U. 1967. 'Beurteilung der glaubhaftigkeit von aussagen.' *Handbuch der psychologie,* 11, 26-181.
Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł. & Polosukhin, I. J. A. i. n. i. p. s. 2017. 'Attention is all you need.' 30.
Vosoughi, S., Roy, D. & Aral, S. 2018. 'The spread of true and false news online.' *Science,* 359:6380, 1146-51.
Wallach, H. M. 2006. 'Topic modeling: beyond bag-of-words.' Paper presented at Proceedings of the 23rd international conference on Machine learning.
Wang, Y., Ma, F., Jin, Z., Yuan, Y., Xun, G., Jha, K., Su, L. & Gao, J. 2018. 'Eann: Event adversarial neural networks for multi-modal fake news detection.' Paper presented at Proceedings of the 24th acm sigkdd international conference on knowledge discovery & data mining.
Williams, T. & Betak, J. 2018. 'A Comparison of LSA and LDA for the Analysis of Railroad Accident Text.' *Procedia computer science,* 130, 98-102.
Wolf, A. & Seebauer, S. 2014. 'Technology adoption of electric bicycles: A survey among early adopters.' *Transportation Research Part A: Policy and Practice,* 69, 196-211.
Yang, F., Mukherjee, A. & Dragut, E. 2017. 'Satirical news detection and analysis using attention mechanism and linguistic features.' *arXiv preprint arXiv:1709.01189*.
Yano, Y., Blandford, D., Maruyama, A. & Nakamura, T. J. B. F. J. 2018. 'Consumer perceptions of fresh leafy vegetables in Japan: An application of word co-occurrence network analysis.'
You, K., Tempo, R. & Qiu, L. J. I. T. o. A. C. 2016. 'Distributed algorithms for computation of centrality measures in complex networks.' 62:5, 2080-94.
Zarocostas, J. 2020. 'How to fight an infodemic.' *The Lancet,* 395:10225, 676.

Zhang, C., Gupta, A., Kauten, C., Deokar, A. V. & Qin, X. 2019. 'Detecting fake news for reducing misinformation risks using analytics approaches.' *European Journal of Operational Research,* 279:3, 1036-52.

Zhang, X. & Ghorbani, A. A. 2020. 'An overview of online fake news: Characterization, detection, and discussion.' *Information Processing & Management,* 57:2, 102025.

Zhang, Z., Feng, G., Xu, J., Zhang, Y., Li, J., Huang, J., Akinwunmi, B., Zhang, C. J., Ming, W.-k. J. J. P. H. & Surveillance 2021. 'The Impact of Public Health Events on COVID-19 Vaccine Hesitancy on Chinese Social Media: National Infoveillance Study.' 7:11, e32936.

Zhou, L., Burgoon, J. K., Nunamaker, J. F. & Twitchell, D. 2004a. 'Automating linguistics-based cues for detecting deception in text-based asynchronous computer-mediated communications.' *Group decision and negotiation,* 13:1, 81-106.

Zhou, L., Burgoon, J. K. & Twitchell, D. P. 2003a. 'A longitudinal analysis of language behavior of deception in e-mail.' Paper presented at International Conference on Intelligence and Security Informatics.

Zhou, L., Burgoon, J. K., Twitchell, D. P., Qin, T. & Nunamaker Jr, J. F. 2004b. 'A comparison of classification methods for predicting deception in computer-mediated communication.' *Journal of Management Information Systems,* 20:4, 139-66.

Zhou, L., Twitchell, D. P., Qin, T., Burgoon, J. K. & Nunamaker, J. F. 2003b. 'An exploratory study into deception detection in text-based computer-mediated communication.' Paper presented at 36th Annual Hawaii International Conference on System Sciences, 2003. Proceedings of the.

Zhou, X., Jain, A., Phoha, V. V. & Zafarani, R. 2020a. 'Fake news early detection: A theory-driven model.' *Digital Threats: Research and Practice,* 1:2, 1-25.

Zhou, X., Wu, J. & Zafarani, R. 2020b. 'SAFE: Similarity-aware multi-modal fake news detection.' *arXiv preprint arXiv:2003.04981*.

Zhou, X. & Zafarani, R. 2018a. 'Fake news: A survey of research, detection methods, and opportunities.' *arXiv preprint arXiv:1812.00315*.

Zhou, X. & Zafarani, R. 2018b. 'Fake news: A survey of research, detection methods, and opportunities.' *arXiv preprint arXiv:1812.00315,* 2.

Zhou, X. & Zafarani, R. 2019. 'Network-based fake news detection: A pattern-driven approach.' *ACM SIGKDD Explorations Newsletter,* 21:2, 48-60.

Zhu, B., Zheng, X., Liu, H., Li, J. & Wang, P. 2020. 'Analysis of spatiotemporal characteristics of big data on social media sentiment with COVID-19 epidemic topics.' *Chaos, Solitons & Fractals,* 140, 110123.

Zuckerman, M., DePaulo, B. M. & Rosenthal, R. 1981. 'Verbal and nonverbal communication of deception.' *Advances in experimental social psychology:* 1-59. Elsevier.