5-2023

Themes and Participants' Role in Online Health Discussion: Evidence From Reddit

Massara Alazazi

The University of Texas Rio Grande Valley

THEMES AND PARTICIPANTS' ROLE IN ONLINE HEALTH DISCUSSION:
EVIDENCE FROM REDDIT

A Dissertation

by

MASSARA ALAZAZI

Submitted in Partial Fulfillment of the
Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Major Subject: Business Administration

The University of Texas Rio Grande Valley
May 2023

THEMES AND PARTICIPANTS' ROLE IN ONLINE HEALTH DISCUSSION:

EVIDENCE FROM REDDIT

A Dissertation
by
MASSARA ALAZAZI

COMMITTEE MEMBERS

Dr. Bin Wang
Chair of Committee

Dr. Murad Moqbel
Committee Member

Dr. Jun Sun
Committee Member

Dr. Diego Escobari
Committee Member

May 2023

ABSTRACT

Alazazi, Massara, Themes and Participants' Role in Online Health Discussion: Evidence from Reddit. Doctor of Philosophy (Ph.D.), May, 2023, 102 pp., 19 tables, 3 figures, references, 166 titles.

Health-related topics are discussed widely on different social networking sites. These discussions and their related aspects can reveal significant insights and patterns that are worth studying and understanding. In this dissertation, we explore the patterns of mandatory and voluntary vaccine online discussions including the topics discussed, the words correlated with each of them, and the sentiment expressed. Moreover, we explore the role opinion leaders play in the health discussion and their impact on participation in a particular discussion. Opinion leaders are determined, and their impact on discussion participation is differentiated based on their different characteristics such as their connections and locations in the social network, their content, and their sentiment. We apply social network analysis, topic modeling, sentiment analysis, machine learning, econometric analysis, and other techniques to analyze the collected data from Reddit. The results of our analyses show that sentiment is an important factor in health discussion, and it varies between different types of discussions. In addition, we identified the main topics discussed for each vaccine. Furthermore, the results of our study found that global opinion leaders have more influence compared to local opinion leaders in elevating the health discussion. Our study has important theoretical and practical implications.

DEDICATION

The completion of my doctoral studies would not have been possible without the love and support of my family. My husband, my mother, my father, my kids, and my siblings, wholeheartedly inspired, motivated, and supported me by all means to accomplish this degree. Thank you for your love and patience.

ACKNOWLEDGMENTS

I will always be grateful to Dr. Bin Wang, chair of my dissertation committee, for all her mentoring and advice. From research funding, research design, and data processing to manuscript editing, she encouraged me to complete this process through her infinite patience and guidance. My thanks go to my dissertation committee members: Dr. Murad Moqbel, Dr. Diego Escobari, and Dr. Jun Sun. Their advice, input, and comments on my dissertation helped to ensure the quality of my intellectual work.

TABLE OF CONTENTS

Page

Page

Page

LIST OF TABLES

LIST OF FIGURES

Page

CHAPTER I

INTRODUCTION

The advancement in information and communication technologies has expedited the social media global penetrance (Puri, Coomes, Haghbayan, & Gunaratne, 2020).Various social media have become popular in engaging users in health related discussion (Park, Conway, & Chen, 2018). Many individuals seek and provide health information and emotional support through online health communities and computer-mediated social support groups (Amith et al., 2020; E. Kim, Scheufele, Han, & Shah, 2017). These platforms provide an anonymous, less stigmatized, and reduced cost and effort medium to exchange health information and initiate health discussions and conversations (L. Chen, Baird, & Straub, 2019). Social networking sites eliminate physical boundaries and allow individuals with diverse health conditions and interests to communicate and share their challenges, insights, and knowledge (L. Chen et al., 2019). Social media content is vibrant and can be reached by large audience and quickly propagate (Puri et al., 2020). Thus, a broad and rich online health content is available for researchers to explore and analyze. Big data capabilities can be leveraged to design better engaging online health communities that facilitate the communication between patients, physicians, and caregivers. As well as, understanding the online health discussion and users interactions helps in improving health awareness (Ayers et al., 2016), combating health-related fake news and rumor

(Pulido, Ruiz-Eugenio, Redondo-Sama, & Villarejo-Carballido, 2020), fostering the adoption of newly generated vaccines and medications, and reduce vaccine hesitancy (Puri et al., 2020).

Due to the importance of vaccines as public health interventions in preventing infectious diseases, mitigating some diseases critical symptoms, and forming herd immunity (Puri et al., 2020), we aim to study online discussion related to two types of vaccines, mandatory and voluntary. People are divided into supporters and anti-vaccines. Our works aims to explore and figure out what topics and themes are discussed on popular social media outlets such as Reddit.

Opinion leaders are vital actors in social media, due to their role in influencing others' attitude and behavior (Mohamad, Ahmad, Salleh, & Sulaiman, 2017; Oueslati, Arrami, Dhouioui, & Massaabi, 2021; Rehman, Jiang, Rehman, Paul, & Sadiq, 2020). They possess social and communication skills. Opinion leaders in online health discussion are found to be active communicators who initiate conversations and provide health advice and information to others (E. Kim et al., 2017). They have the ability to persuade others with their view of point and inspire them (Mohamad et al., 2017). As a result, we aim to study the impact of opinion leaders' participation on online health discussions. Furthermore, identifying opinion leaders is of high importance as they play a critical role in the diffusion of information in their social networks (Rehman et al., 2020).

The research questions that we address are:

1. What are the differences between mandatory and voluntary vaccine online discussions in terms of the topics discussed, the sentiment expressed, and the terms associated with each of them?

2. How different types of opinion leaders' affect the dynamics of online health conversation including post volume and responsiveness?

1.1 Abstracts of Two Essays

To answer the above research questions, this dissertation is organized into two essays. The first essay involves a comparison of online discussions of mandatory versus voluntary vaccines using text mining. The second essay involves examining the impacts of opinion leaders on online health-related discussions. The abstracts of the two essays are as follows:

1.1.1 Essay 1: Comparison of Voluntary versus Mandatory Vaccine Discussions in Online Health Communities: A Text Analytics Approach

Vaccines are vital health interventions. However, they are controversial, and some people support them while others reject them. Social media discussion and big data are a rich source to understand people's insights about different vaccines and the related topics that concern most of them. This study aims to explore the online discussions about mandatory and voluntary vaccines using text analysis techniques. Reddit social platform is popular in online health discussion and thus data from Reddit is analyzed. The results show that different aspects are discussed for different types of vaccines. The discussion of mandatory vaccines is more interactive and is focused on the risks associated with them. Voluntary vaccines' discussion is focused on their effectiveness and whether to get them or not. The study has important implications for health agencies and researchers as well as for healthcare providers and caregivers.

1.1.2 Essay 2: The Role of Opinion Leaders in Elevating Online Health Discussion: Evidence from Reddit

Online health discussion has grown dramatically. Different health conditions are discussed on social networking sites interactively. As conversation participants vary in their roles and activities, opinion leaders are vital actors in these online discussions and have an influential impact on others' attitudes and behavior. This study aims to examine opinion leaders' impact on online health discussion dynamics and participation. Opinion leaders are classified into different categories according to their content, influence, and participation. The study used data from Reddit platform to test the impacts of different types of opinion leadership and related hypotheses. While Reddit is the sixth most visited website in the United States, its dynamics are not sufficiently examined. Data about health discussion has been analyzed using social network analysis, text analytics, econometric models, and machine learning to identify opinion leaders and their impacts on other participants' involvement in the discussion. Understanding these dynamics helps in better designing the platform to fulfill the needs of its users and employ the right type of opinion leaders to disseminate health information needed and combat health misinformation on social media.

CHAPTER II

COMPARISON OF VOLUNTARY VERSUS MANDATORY VACCINE DISCUSSIONS IN ONLINE HEALTH COMMUNITIES: A TEXT ANALYTICS APPROACH

2.1 Introduction

Vaccination is an essential part of individuals' health interventions. They are found to contribute to the reduction of mortality rates (Aghili & Lapointe, 2019). However, many individuals either refuse vaccinations or doubt their effectiveness and efficiency (Glanz et al., 2017). Statistics show that between 10 to 20% of parents refuse or postpone at least one vaccine for their children (Daley, Narwaney, Shoup, Wagner, & Glanz, 2018). Because some vaccinations are controversial, many people turn to social media outlets to get information about vaccines (Daley et al., 2018; Love, Himelboim, Holton, & Stewart, 2013) or to influence others by disseminating their beliefs and theories (Bello-Orgaz, Hernandez-Castro, & Camacho, 2017). Online blogs, microblogs, discussion boards, videos and their content can impact the decision of vaccination for many individuals (Love et al., 2013). Particularly, new parents could be affected by their social networks regarding their children's vaccination due to the lack of experience, hesitancy, and others (Brunson, 2013).

Many debates are available on social media between supporters and rejecters of vaccines (Massey et al., 2020). It is vital to explore this content to understand the accuracy of information, the beliefs and attitudes of people, their concerns, and the topics that dominate the online discussion. Vaccines differ in their deliberative nature. Some vaccines have more agreement by individuals and science regarding their necessity and thus are mandatory. However, other vaccines such as the flu vaccine are more controversial. In addition, due to governmental restrictions and policies, some vaccines are mandatory, and others are voluntary. When a vaccine is voluntary, people have less motivation to be vaccinated (Fukuda et al., 2014).

This study aims to compare and contrast online discussions about mandatory versus voluntary vaccines. Previous research has focused on studying one type of vaccine or studying vaccines in general. However, knowing people's insights and opinions about different types of vaccines help in shaping health policies and determining the need to convert some vaccines from voluntary to mandatory if necessary. In addition, our data analysis is based on an online social platform – Reddit – that is not adequately explored, unlike Facebook and Twitter. Exploring various online platforms' content can help in providing us with a better understanding of people's conversations and their associated feelings, attitudes, and topics of concern. Our study uses text analytics to analyze the terms that are correlated with each vaccine discussion, the affect and sentiment expressed, and the topics discussed for each type of vaccine. The research questions that this research addresses are:

- What are the differences between mandatory and voluntary vaccine online discussions in terms of the topics discussed, the sentiment expressed, and the terms associated with each of them?

- How do text mining results inform us on the participation in voluntary vs. mandatory vaccine discussions?

We chose the flu vaccine as the voluntary vaccine to study and the measles, mumps and rubella (MMR) vaccine as the mandatory one. Our research has two major findings and contribution to the literature and health prevention practice. First, our results show the differences in online discussions of mandatory vs. voluntary vaccines. Flu vaccine online discussions have more emotional expressions compared with the MMR vaccine discussion. In addition, flu vaccine discussions have dominant topics such as the effectiveness of the vaccine and getting the vaccine annually. The topics that dominate the MMR vaccine discussions include the debate about the MMR vaccine and the risks associated with it in addition to the need for evidence about these risks. Second, our results identify sentiment, vaccine discussion topic, and length of thread initial post as important factors that affect online vaccine discussion participation.

2.2 Literature Review

2.2.1 Online Health Communities (OHCs) Participation

Social media is leveraged by many patients, physicians, and caregivers to generate health related-content (Hajli, Sims, Featherman, & Love, 2015). Online health communities are widely used due to the convenience, anonymity, and support provided (Alazazi & Ayaburi, 2019). Many health issues are discussed on social media and their related content is studied including eating disorders (Sowles et al., 2018), smoking (Myneni, Fujimoto, Cobb, & Cohen, 2015), drugs (Mukherjee, Weikum, & Danescu-Niculescu-Mizil, 2014), breast cancer (Elhadad, Zhang, Driscoll, & Brody, 2014), mental health (Shepherd, Sanders, Doyle, & Shaw, 2015), chronic diseases (Q. B. Liu, Liu, & Guo, 2020) and many others. Participation in health communities and forums is motivated by many factors including reciprocity, altruism, and homophily (Alazazi & Ayaburi, 2019; X. Zhang, Liu, Deng, & Chen, 2017). Patients refuge to the online environment to fulfill their needs such as sharing their experiences, managing their illness, learning from similar others, or getting others' support, opinions, or advice (Hajli et al., 2015; Huh, McDonald, Hartzler, & Pratt, 2013). Social networking sites and blogs give users the ease of asking health-related questions, sharing their thoughts, watching educational videos related to their health condition, and reading peers' and physicians' posts and replies (Hajli et al., 2015). Communication with physicians on these social sites may lead to both better self-health management and better patient-physician relationship (Q. B. Liu et al., 2020). Even though online health communities and forums are frequented by many individuals, the credibility of health and medical information generated by users is a concern (Mukherjee et al., 2014). If patients take serious health actions based on the information they receive from online content, the outcome may not be desirable if the information is inaccurate.

Participation in online health communities including posting, replying, voting on a post, and others has an impact on users' feelings, perceptions, and even behavior (Willis, 2018). For instance, the influence of user-generated content in online health communities on health behavior was studied (Willis, 2018). The study focused on the online user-generated content exchange between patients with chronic diseases and medication adherence. The discussion was mainly concentrated on three themes: "striving for pain relief; negotiating potential side effects, and finding the new normal" (Willis, 2018). In addition, participation in OHCs is affected by the relationships formed within the community and the information exchanged (L. Chen et al., 2019). The social capital of participants impacts social support exchange within the community.

And the social support exchange helps in improving health knowledge and literacy. Social support with the community could have different forms including providing informational support, seeking informational support, providing emotional support, seeking emotional support, and companionship (Xi Wang, Zhao, & Street, 2017). Furthermore, some factors impact the different knowledge-sharing behavior and participation in the OHC. These factors include self-worth, perceived social support, reputation, face concern, and sharing costs (Yan, Wang, Chen, & Zhang, 2016)

2.2.2 Mandatory vs Voluntary Vaccines

Vaccination is a vital tool to prevent many illnesses (Haverkate et al., 2012). The occurrence of many diseases has declined after the implementation of their vaccines such as the measles, mumps, and rubella MMR. However, there is a wide spectrum of vaccines available. Some vaccines are mandatory, and others are voluntary in the US. Some vaccines are required before a child enters school. Other vaccines are voluntary and depend on the individual or on the parents' preference. Voluntary vaccines decision depends on personal social norms (Gesser-Edelsburg, Walter, Shir-Raz, & Green, 2015), religious issues (Galanakis, Jansen, Lopalco, & Giesecke, 2013), beliefs, risk of the illness, or the vaccine effectiveness and side effects (F. H. Chen, 2006; Fukuda et al., 2014). Previous research proposed a risk assessment model to vaccination decision-making by comparing the payoff and the risk. By modeling a vaccination game, the study found that social networks of individuals play a major role in the decision of vaccination (Fukuda et al., 2014). Another study focused on the controversial debate about mandatory vs voluntary healthcare workers vaccination (Galanakis et al., 2013). There is a need to protect them and reduce the spread of infectious diseases. On the other hand, health ethics and individuals' autonomy need to be considered. The study concludes that there is a moral need for vaccination of healthcare workers and enforcing vaccination would be more fruitful (Galanakis et al., 2013).

Online health communities play a key role in discussing vaccine controversies. A recent study explored their role in terms of knowledge delivery for pro and anti-vaccine movements (Aghili & Lapointe, 2019). The study found that both knowledge delivery practices online and offline impact each other while differ in their configurations. The study compared the format of the knowledge, the delivery of the knowledge, and the availability of the knowledge. Social media has an impact on individuals' decision of vaccination (Langley, Wijn, Epskamp, & Van Bork, 2015; Margolis, Brewer, Shah, Calo, & Gilkey, 2019). For instance, an online experimental design targeted parents of girls to test the impact of online information about HPV vaccine on them from the lens of the health belief model (Langley et al., 2015)s. The study used a novel exploratory network analysis and found that perceived efficacy is influential while cues to action is not. Moreover, a previous study applied the semantic network analysis on articles shared on Twitter about vaccines and analyzed the resulting networks. The study also identified positive, negative, and neutral sentiment in the text of the shared articles (Kang et al., 2017). In their study the researchers used manual sentiment coding of the articles.

2.2.3 Text Mining

Big data and the enormous amount of unstructured texts available on social media encouraged researchers to analyze these texts, extract knowledge and patterns, and infer results by applying techniques such as classification, clustering, machine learning, and social network analysis (Salloum, Al-Emran, & Shaalan, 2017). Text mining helps in providing an idea or overview of unstructured text information in the online environment (J. Kim, Bae, & Hastak, 2018). The most frequently used words could be extracted and analyzed to get data trends and patterns on social media in different disciplines and topics. In addition, the association between the frequent words shows the correlation between terms and which terms social media users use together more frequently (J. Kim et al., 2018).

Previous research focused on Twitter and Facebook extracted text. Different text analytics techniques are used to provide textual orders and thus provide key themes in data (Salloum, Al-Emran, Monem, & Shaalan, 2017). Sentiment analysis is a subfield of natural language processing that is used to extract sentiment-related words, emotions and find their polarity toward a specific topic, product, organization, or other entities (Yue, Chen, Li, Zuo, & Yin, 2019). Previous research used sentiment analysis using both supervised and unsupervised machine learning techniques. For example, one study analyzed tweets from Twitter about a particular product or movie based on sentence-level sentiment identification using the SentiWordNet software (Rout et al., 2018). The study found that tweeters use hashtags to express their emotions. Sentiment analysis helps e-commerce platforms to analyze their products and services and discover customers' preferences (Aghili & Lapointe, 2019; Yue et al., 2019). For instance, a study used emotional text mining to analyze customer profiling for brand management of sportswear from Twitter. Cluster analysis with a bisecting k-means algorithm and correspondence analysis are used to analyze the data. The study identified Twitter users' symbolic categories, measured their sentiment, and found their representations of the sportswear brand (Rout et al., 2018). The sentiment expressed in online posts has an impact on the health support group participation (Zheng, Li, & Farzan, 2018). One study focused on Facebook support groups and found that posts with positive sentiment generated a generally positive tone in the comments (Zheng et al., 2018). For posts with negative sentiment, the results were mixed: some had comments with positive tones, some comments had negative sentiment, and others with a mixture of positive and negative sentiment. Furthermore, negative posts generated a larger number of participation (comments) compared with positive posts. This variation in participation emphasizes the supportive nature of these online groups. Users tried to cheer up and support people with negative posts either by providing positive responses or sharing their negative experiences with them. According to a study of Twitter tweets sentiment, users express positive and negative sentiment when engaging in an actual conversation, while users retweet posts with neutral to negative sentiment (Hemmings-Jarrett, Jarrett, & Blake, 2018). The study emphasized the importance of differentiating between user groups based on their participation and interactions.

Another study analyzed news channels' online textual data from Facebook using the RapidMiner tool. The results indicate that the most covered topics on these news channels pages were US election news. CNN had the most shared posts about the topic (Mhamdi, Al-Emran, & Salloum, 2018). Another research paper proposed a deception detection mechanism for crowdfunding projects by considering both static project information and dynamic communication between funders and fundraisers for classification (Siering, Koch, & Deokar, 2016). Cues extracted from the text such as content-based cues and linguistic cues to detect fraudulent crowdfunding projects are used to detect fraud using machine learning algorithms.

Text mining for health-related content on social media has also been applied in previous studies (Wimmer, Yoon, & Sugumaran, 2016). One study analyzed flu-related tweets on Twitter using network analysis. The study found that effective information about flu is generated by accounts found in the important Twitter accounts and these tweets would stay active for a longer period compared with other individual accounts' tweets (Yun et al., 2016). Another study investigated the opinions of Twitter users about the influenza vaccine using natural language classifiers to identify vaccine attitudes and behaviors (Huang et al., 2017). Using the MedHelp.org health platform, a study performed text mining to identify the main stakeholders participating in lung cancer, breast cancer, and diabetes forums. Patients and caregivers were the main population of participants while specialists formed the minority of them (Lu, Wu, Liu, Li, & Zhang, 2017). The discussed topics included symptoms, drugs, procedures, examinations, and complications. Furthermore, sentiment analysis was performed using sentiment lexicon software SentiWordNet. Cluster techniques such as topic and probabilistic clustering were applied as well as keyword extraction and topic identification. An interesting study explored the number of messages on different social media platforms during the measles outbreak in the Netherlands in 2013 and compared them with both the number of related-online news and the number of reported measles cases (Mollema et al., 2015). Classification, text mining techniques, and manual sentiment analysis were used. However, in regards to the application of sentiment analysis, the health domain was found to be behind when compared with other domains (Zunic, Corcoran, & Spasic, 2020). The performance of methods used to measure sentiment for health is lower than in other disciplines. As a result, more research is needed to address this deficiency in health-related content. Furthermore, Covid-19 vaccine attitudes on social media such as twitter were analyzed, and sentiment analysis was applied to determine the vaccine perceptions of Indian citizens. Topic modeling was also performed to reveal the general topics of Covid-19 vaccine discussed by the pubic on social media (Praveen, Ittamalla, & Deepak, 2021). The study found that most of the discussion had a neutral tone and the main concern discussed was fear of allergic reactions of the vaccine. Covid-vaccine discussion on Reddit was explored using sentiment analysis and topic modeling and found that the more positive sentiment was expressed than negative sentiment and the most discussed topic was the side effects of the vaccine (Melton, Olusanya, Ammar, & Shaban-Nejad, 2021).

2.3 Theory and Hypothesis Development

2.3.1 Message Framing Theory

The way a message is presented to the audience has a great impact on how people process its content and react to it (Smith & Petty, 1996). The framing postulate of the prospect theory suggests that individuals evaluate information about uncertain alternatives in two ways. They frame messages either by using positive framing that focuses on potential gains or by using negative framing focusing on potential risks or losses (Smith & Petty, 1996). Furthermore, negatively framed information might have different judgments by participants compared with positively framed messages. Individuals' preference for an option depends on whether its positively or negatively framed (Finney & Iannotti, 2002). Health-related messages can be framed to show the benefits of a health-related action or show the negative consequences of unhealthy behavior or action (Van't Riet et al., 2016). The framing of health-related messages influences others' health perceptions and behavior (Rothman & Salovey, 1997). In other words, the words, expressions, and emotions used in a message impact the responses to this message (Finney & Iannotti, 2002). However, the context of health-related message framing plays a role in its effectiveness and needs to be examined (Gallagher & Updegraff, 2012).

In the context of vaccine discussion and message framing on social media, the way a message is framed impacts people's responses to it. Vaccines like other medical interventions have potential risks and benefits. Some messages are expressed in a negative way showing the negative consequences and risks associated with a vaccine or negative personal experience, while other messages are framed in a positive way describing the benefits of getting a vaccine such as preventing disease complications and reducing disease spread. Social media users have various insights and opinions about vaccines. The synthesis and composition of their posts vary too. Some of them would use lengthy informational posts, while others use abbreviated more emotional posts that could have negative or positive emotions regarding the vaccine. This variation in message framing may impact the related responses by other social media users. Many people turn to social media to get health-related information. Thus, they might look for information that benefits them and interact with them until their concerns are addressed. Other people are more emotional and discuss negative side effects of vaccines that they heard about or encountered. Or they might discuss the positive experience they had when getting a vaccine. Emotions play a role in information sharing behavior in social media (Brady, Wills, Jost, Tucker, & Van Bavel, 2017; Chawla & Mehrotra, 2021; Stieglitz & Dang-Xuan, 2013). In OHCs people seek emotional support from others, and thus using positive emotions in a post is considered beneficial to others and may motivate them to participate in the discussion (K. Zhao et al., 2014). Positive emotions in health-related messages reflect empathy and reassurance (E. Kim, Hou, Han, & Himelboim, 2016). Previous research has found that positive affect expressed in online messages reflects a sense of community and motivates participation continuity (Joyce & Kraut, 2006). Emotions are found to increase attention and arousal. When users read posts that use positive sentiment, some of their needs are fulfilled and their cognitive involvement may increase (Stieglitz & Dang-Xuan, 2013). In addition, reading posts with positive emotions boost others' mood and positive feelings (E. Kim et al., 2016).Thus, they tend to have behavioral responses such as participating and joining the discussion. Thus, we hypothesize:

Hypothesis 1a: Positive emotions expressed in a vaccine-related initial post will be positively associated with vaccine thread participation.

Both negative and positive emotions are found to attract others' attention (Stieglitz & Dang-Xuan, 2013). Emotions are found to influence user evaluation of content and their behavior (Xiaohui Wang & Lee, 2020). However, negative emotions in social media posts have conflicting impacts in the literature (Keib et al., 2018). One study found that short texts such as tweets with negative sentiment spread faster than messages with neutral or positive sentiment (Ferrara & Yang, 2015). However, different events may stimulate different sentiment patterns with highly anticipated events leading to positive sentiment, while unexpected sudden events like disasters and emergencies generating more negative sentiment (Ferrara & Yang, 2015). Another study revealed that the context matters when it comes to the relationship between sentiment and social media response (Brady et al., 2017). Positive moral emotions in tweets related to same sex marriage were retweeted more than those with negative moral emotions, while the opposite was true for tweets related to climate change. For gun control, tweets with negative and positive moral emotions were retweeted and diffused at almost the same rate.

Empirical research on the diffusion of health-related messages on social media generally shows a positive relationship between sentiment and sharing behavior. For example, positive messages related to smoking are shared more on social media (H. S. Kim, Lee, Cappella, Vera, & Emery, 2013). Additionally, cancer-related tweets with hope and no fear emotions expressed are associated with higher retweet activities (Xiaohui Wang, Chen, Shi, & Peng, 2019). Fear emotions, which are considered a part of negative emotions, had a negative impact on the virality of the tweet. Similarly, initial posts of blogs with negative sentiment on Steemlt, a Reddit-like social site, were associated with less steem dollar earned (Thelwall, 2018). In our research context, vaccine-related posts with negative emotions increase others' vaccine hesitancy and fears and discourage users to participate in the discussion to maintain their positive image. Hence, we suggest that vaccine posts with a negative sentiment would have less participation by other users.

Hypothesis 1b: Negative emotions expressed in a vaccine-related initial post will be negatively associated with vaccine thread participation.

Furthermore, the content of the initial post is important. The main topic or the theme discussed in the post could be of high or low interest to the audience (Xiaohui Wang et al., 2019). Some topics could be more attractive to social media users than others (Monselise, Chang, Ferreira, Yang, & Yang, 2021; Xiaohui Wang & Lee, 2020). For instance, cancer-related tweets with social themes such as personal struggle stories lead to more diffusion (Xiaohui Wang & Lee, 2020). Another study about health-related conversations on Twitter examined the themes discussed by different participants (W. W. Xu, Chiu, Chen, & Mukherjee, 2015). The salient themes of discussion included health knowledge sharing theme (seeking and providing health information and experience), activism, advocacy, and promotion theme that is concerned with raising health awareness, and community building theme. The study found that the health knowledge sharing theme was the most frequent theme with 33% of the entire conversation. For vaccines-related discussion, users would be more interested in some themes and topics than others such as discussing the benefits and risks of vaccines (Shoup et al., 2015). Thus, we hypothesize:

Hypothesis 2: Different vaccine topics discussed in the vaccine-related initial post will have different influences on vaccine thread participation.

Posts differ in their length. Some users compose very abbreviated posts with few words, while others write in detail the issue they are discussing and explain their points elaborately. Due to the high volume of users and posts on social media, shorter posts were found to become more preferred by users on SteemIt as they can be read and accessed more quickly (Thelwall, 2018). For health-related discussion, users usually have some concerns or questions about a specific health issue or treatment. As a result, they tend to ask questions. Other users can answer the questions according to their knowledge or experience and participate in the thread. Others might tell a lengthy story explaining their struggle or journey with the health issue. This variation would have an impact on users' interaction with the post. Shorter posts on Reddit depression community are found to be related to offering advice, providing support to others, or related to therapy (Feldhege, Moessner, & Bauer, 2020). Moreover, for vaccines-related discussion, many people have vaccine hesitancy and would ask others about information related to vaccines. Asking questions serves as an alert that encourages other members to answer and share their knowledge about the asked issue and benefit others. In addition, posts with questions usually have less words than story telling posts. With the increased number of available posts on social media outlets, people tend to read and interact more with shorter posts. Thus, we hypothesize:

Hypothesis 3: The length of a vaccine-related initial post will be negatively associated with the vaccine thread participation.

High frequency words in the first post of a thread on Steemlt were associated with increased value to others and earned more value compared with posts with rare words (Thelwall, 2018). In the vaccine's context, some words are more related to a vaccine than others. For example, when discussing the flu vaccine, people may talk more about the season of the flu and the time to receive it as it is a seasonal vaccine. In contrast, when discussing the MMR vaccine, individuals might talk about the age a child should take it. High-correlated words with a vaccine are the words that are found more in the same discussion. Using such terms in an initial post might interest the readers and get their attention. The users may feel more connected with the discussion and thus participate more in the thread. Thus, we hypothesize:

Hypothesis 4: The number of high-correlated words used in the vaccine-related initial post will be positively associated with the thread participation.

The proposed model is summarized in Figure 1.

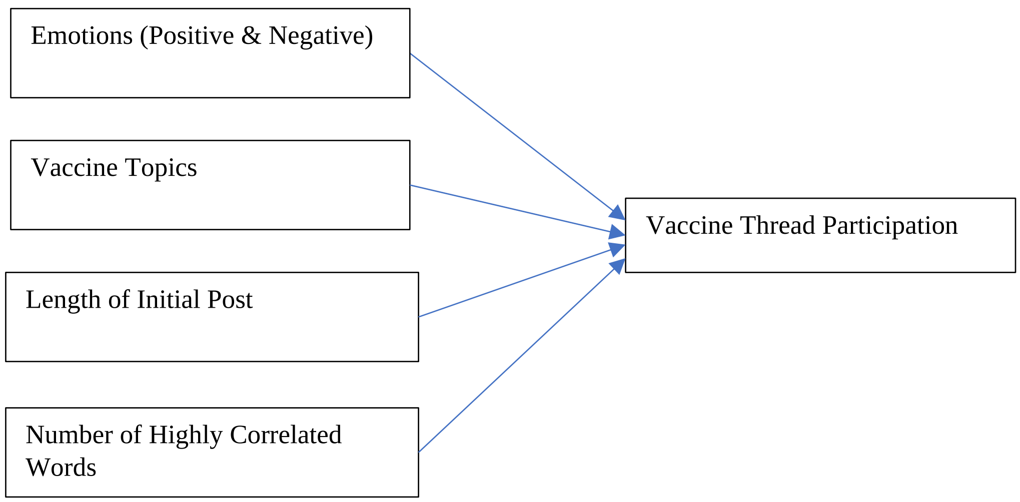

Figure 1. The proposed model

2.4 Institutional Background

As one of the largest online communities (Singer, Flöck, Meinhart, Zeitfogel, & Strohmaier, 2014), Reddit is a social networking site that encompasses a variety of online forums and communities and was dubbed “the front page of the internet” (Singer et al., 2014). In 2018, Reddit was ranked as the fourth most visited social website in the United States and the sixth worldwide with 234 million users (Delnevo et al., 2018). In 2020, its user base has reached 430 million active monthly users and 130,000 active communities according to statistics on its website. Reddit is broken down into communities called “subreddits” that focus on different topics such as news, politics, gaming, and videos. The platform provides interactive features to engage the users. For instance, users can signal their support of a post by clicking on the upvote arrow or click on the downvote arrow to indicate their disapproval. The total score of the post is shown to indicate the number of upvotes minus the number of downvotes.

Academic research has explored the content of Reddit discussion. For example, two studies investigated the text polarity, age, geographic distribution of users, and product acquisition related to discussion on e-cigarette use (Brett et al., 2019; Zhan, Zhang, Okamoto, Zeng, & Leischow, 2019). Another study explored the questions posted on Reddit about gout illness and classified them into 13 categories such as symptom uncertainty and diagnosis (Derksen, Serlachius, Petrie, & Dalbeth, 2017). Another research examined user posting behavior on Reddit based on network structure and revealed that most users participate in one Reddit community (Buntain & Golbeck, 2014). Another study on the weight loss subreddit examined how online interactions affect weight loss in regards to the number of votes and replies received and used topic modeling and hierarchical clustering algorithm to identify global topics and local clusters (Yang Liu & Yin, 2020). Reddit health communities are growing and Reddit has shown a 43% increment in its health and fitness content (Cassis, 2019).

2.5 Sample

We collected publicly available data about one mandatory vaccine – the MMR vaccine – and one voluntary vaccine – the flu vaccine – from Reddit. The MMR vaccine is a vaccine against three viruses: Measles, Mumps, and Rubella. We chose the MMR vaccine as the mandatory vaccine of our research for two reasons. First, the three diseases that the MMR vaccine protects humans against can cause serious health complications including bronchopneumonia, brain damage, mental retardation, fetal anomalies, and parotitis (Watson, Hadler, Dykewicz, Reef, & Phillips, 1998). Hence, the MMR vaccine is vital. Second, the MMR vaccine had an unproven linkage with autism that created a controversy (Watson et al., 1998) and there has been debate on whether children should receive the vaccine. As a result, it is important to study this vaccine, due to its importance but possible side effect. We chose the flu vaccine as the voluntary vaccine to examine. While it is optional to take the flu vaccine, it has many benefits according to the Centers for Disease Control and Prevention (CDC) including the reduction of risks associated with the flu illness and reducing flu-related doctor visits by 40-60% (CDC, 2020). On the other hand, not all people agree with these benefits. Many claim getting the flu after receiving the vaccine and the vaccine make them sick (Nyhan & Reifler, 2015). In addition, there are potential sources of bias in studies related to flu vaccine effectiveness, particularly among the elderly (Trucchi, Paganino, Orsi, De Florentiis, & Ansaldi, 2015). In addition, mercury in thimerosal found in some flu vaccines could be decomposed to toxic compounds, making it of high concern for many parents and individuals (Drum, 2009). Hence, the benefits as well as potential side effects of both vaccines allow us to examine how online discussion differs on mandatory versus optional vaccines.

Initial thread posts that were available on Reddit in May 2020 with “MMR vaccine” or “measles, mumps, and rubella vaccine” in the title were collected for the MMR vaccine and initial thread posts containing “flu vaccine”, “influenza vaccine” in the title were collected for the flu vaccine. Next, we collected all replies to these initial thread posts. We obtained a total of 11,176 posts on the MMR vaccine and 10,152 posts on the flu vaccine. The descriptive statistics for the two types of vaccine posts and comments are shown in Table 1. There were more comments on MMR vaccine-related posts compared with the flu vaccine discussion. The post score is a rating Reddit gave to posts based on factors including up votes and down votes. Up votes are given by users to posts that they think that posts contribute to a conversation, and down votes are given to posts that user thinks they do not contribute to the conversation in the subreddit. Posts about MMR vaccines had higher average score than flu vaccine posts.

Table 1. Sample descriptive statistics

Variable	**Min**	**Max**	**Mean**	**Std. Dev.**
Flu Vaccine Posts (N1=10,152)				
# Comments on a Post	1	1,982	71.27	246.53
Post Score	0	49,292	1,146.84	5,552.27
Up Votes (%)	0.14	1	0.81	0.18
MMR Vaccine Posts (N2=11,176)				
# Comments on a Post	1	3,468	61.21	279.31
Post Score	0	77,348	881.47	6,159.56
Up Votes (%)	0.18	1	0.83	0.17

2.6 Text Analytics Results

Our text mining process includes data collection, data preprocessing and cleansing, and text analytics. After collecting the posts, we first performed text preprocessing to improve the effectiveness of text analytics by removing unnecessary text from the analysis that may overwhelm the analysis (Igawa, Almeida, & Zarpelão, 2015). This process included converting the text to lower case, removing stop words that do not provide useful information (Kühl, Mühlthaler, & Goutier, 2018), and reducing the words to their stems (Vijayarani, Ilamathi, & Nithya, 2015). In addition, we removed punctuations and numbers and stripped extra white space in the text. To keep the amount of text we analyze manageable, we only retained terms that appeared in at least 2.5% of the posts. We generated our document term matrix using these remaining words and performed our subsequent term association and topic modeling analyses.

2.6.1 Term Association Analysis of All Posts

For the term association analysis on the flu vaccine-related posts, we started with the term ‘flu’ and identified 18 first-level terms that had at least a 0.20 correlation with flu. Then we identified 76 second-level terms that had at least a 0.20 correlation with one of the first-level terms. Due to space constraint, we summarize only the first-level terms and their correlations with the term ‘flu’ in Table 2. Next, we plot the resulting first- and second-level term associations as edges in a term network in Figure 2. The size of each vertex is proportional to its eigenvector centrality in the network. That is, the larger the node representing the term, the more influential the term is in the network. The network graph reveals that the most important terms in flu vaccine-related discussions were vaccine, flu, year, get, virus, influenza, strain, risk, effect, people, even and got. These term association results reveal that flu vaccine-related discussions focused on the requirement of annual vaccination, different strains of the flu viruses, protection offered by the flu vaccines, the impact of the vaccine on the immune system, and the risks involved in getting the flu shots.

Table 2. Flu vaccine posts term association analysis results

Term 1	Term 2	Correlation
flu	shot	0.53
flu	get	0.46
flu	vaccin	0.42
flu	year	0.37
flu	peopl	0.28
flu	everi	0.27
flu	strain	0.26
flu	got	0.25

Table 2, cont.

flu	immun	0.24
flu	season	0.23
flu	week	0.22
flu	effect	0.21
flu	help	0.21
flu	still	0.21
flu	virus	0.21
flu	even	0.20
flu	like	0.20
flu	risk	0.20

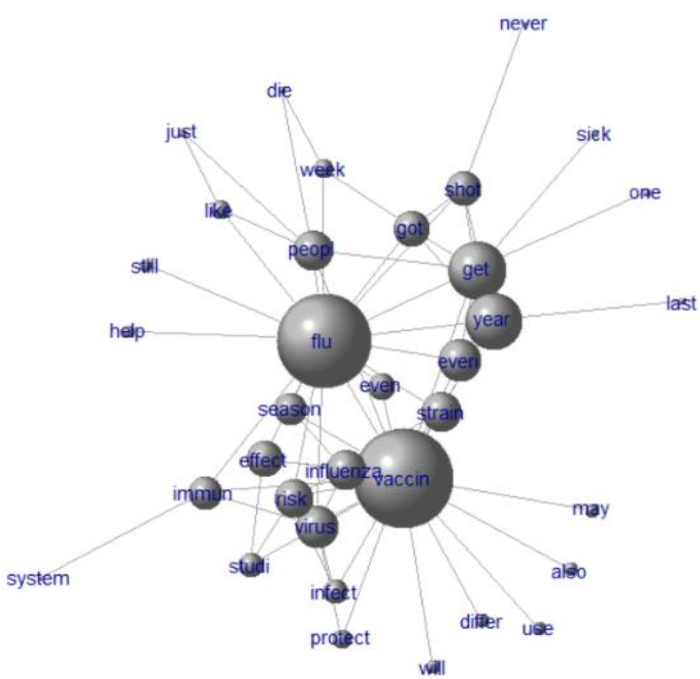

Figure 2. Term association network for flu vaccine-related posts

We performed a similar term association analysis for MMR vaccine-related posts using 'mmr' as the starting term and identified six first-level terms with at least a 0.20 correlation with 'mmr' and 78 second-level terms with at least a 0.20 correlation with one of the first-level terms. We summarize the first-level terms and their correlations with 'mmr' in Table 3. Figure 3 shows the term association network graph for the MMR vaccine-related posts. These results suggest that the discussions on the MMR vaccines focused on parents vaccinating their children and the risks associated with the vaccine. The linkage between the MMR vaccine and autism and the need for studies and evidence to support or negates this issue were also discussed. The term network diagram for the MMR vaccine discussion is denser.

Table 3. MMR vaccine posts term association analysis results.

Term 1	Term 2	Correlation
mmr	vaccin	0.34
mmr	autism	0.23
mmr	risk	0.22
mmr	measl	0.22
mmr	studi	0.22
mmr	caus	0.20

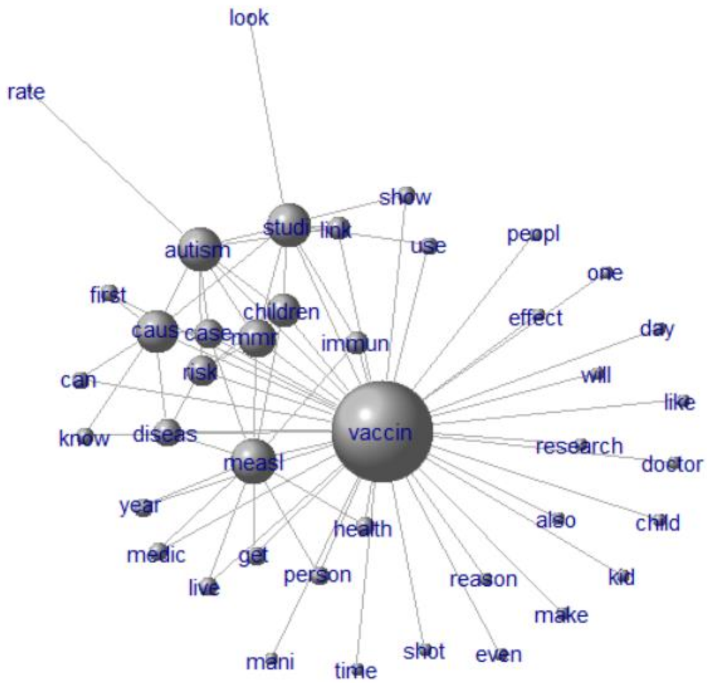

Figure 3. Term association network for MMR vaccine-related posts.

We summarize the comparison of the term association network analysis results for the discussions on the two vaccines in Table 4. We note that flu vaccine discussions mentioned more terms or vertices with more term associations or edges higher than 0.20 than the MMR vaccine-related discussions. However, the term association network density is higher for the MMR vaccine-related discussions. These results suggest that the discussions on the flu vaccines are more sparse compared with the MMR vaccine-related discussions.

Table 4. Comparison of term association network analysis results

Characteristic	Flu Vaccine Posts	MMR Vaccine Posts
# Vertices	131	97
# Edges	94	84
Network Density	0.013	0.019

2.6.2 Sentiment Analysis Results of All Posts

We next performed sentiment analysis (Biswas, Mukhopadhyay, & Gupta, 2018). We analyzed the emotions expressed in each vaccine discussion of all posts using the Linguistic Inquiry and Word Count (LIWC) 2015 software (Pennebaker, Booth, Boyd, & Francis, 2015). The LIWC has an internal dictionary that defines the words to be counted in the required text file (Biswas, Sengupta, & Chatterjee, 2020). The LIWC2015 dictionary encompasses 6400 word stems, words, and emoticons. Sub dictionaries and word categories are defined using words entries included in dictionaries. Scales of categories are defined by a list of dictionary words. Some categories follow a hierarchical arrangement, which means that some categories are included in a broader category. For example, sadness words are also included in the broader negative emotions category. The dictionary also captures stems of the words in order to group words with the same stem together. There are main steps in the creation of the main LIWC2015 dictionary including word collection, judge rating, base rate analysis, candidate word list generation, psychometric evaluation, refinement, and addition of summary variables including analytical thinking, clout, authenticity, and emotional tone (James W. Pennebaker & Kayla Jordan, 2015).

Examples of flu-vaccine discussion with positive sentiment are “Thank you! Vaccines are my passion and I'm working hard in staying in this field for the rest of my career” and “I got my flu shot! ... It took less than five minutes for it all and I’m SO glad I got it. I can’t wait to not get the flu!” Examples of flu-vaccine discussion with negative sentiment are “This is bad information. A healthy adult still needs the flu shot…” and “people who won't get the flu shot (or take proper precautions) make me so angry.”

Examples of MMR vaccine discussion with positive sentiment are: *"Today my daughter was in the lucky 5% of people who developed a mild rash and fever as a result of the vaccine. Lucky because I'd much rather her have a mild fever and some polkadots than fucking measles"* and *"[I]t's great to see Vaccine rates have increased during this measles outbreak we need herd immunity!"* Examples of MMR vaccine discussion with negative sentiment are: *"... I'm up to date on all my vaccines, but my titres for MMR show I keep losing immunity ... I'm angry that we're not doing more to stop this pro-plague idiocy. Get vaccinated!"* and *"[f]riend of mine's kid got roseola a few days after the MMR. Blames all vaccines, again. Purple made me angry."*

Table 5 summarizes the LIWC sentiment analysis results including positive emotions, negative emotions, anxiety, anger, and sadness. Anxiety, anger and sadness are considered subcategories of negative emotions. Overall, the results suggest that discussions on flu and MMR vaccines were not statistically different in terms of positive emotions but the discussions on the flu vaccine were more negative compared with those on the MMR vaccine. A more in-depth analysis on the negative emotions shows that there were also more negative emotions of anxiety and anger in the flu vaccine-related posts than the MMR vaccine-related ones. However, discussions on the two types of vaccines do not differ significantly in the negative emotion of sadness.

Table 5. Comparison of sentiment analysis results of all posts

Sentiment Measure	**Mean of Flu Vaccine Posts**	**Mean of MMR Vaccine Posts**	**t-Stat on Difference in Means**
Positive emotions	3.570 (0.082)	3.372 (0.078)	1.570
Negative emotions	2.712 (0.056)	2.415 (0.050)	4.372***
Anxiety	0.428 (0.017)	0.317 (0.017)	4.196***
Anger	1.125 (0.039)	0.962 (0.033)	3.501***
Sadness	0.306 (0.016)	0.285 (0.017)	0.888

Table 5, cont.

Notes: Standard deviations in parentheses. *** p<0.01; ** p<0.05; * p<0.10.

2.6.3 Topic Modeling Analysis Results of All Posts

We first performed topic modeling of the posts using Latent Dirichlet Allocation (LDA). LDA is a text mining technique that helps in finding hidden relationships among text documents and discover topics among them (Jelodar et al., 2019). The use of topic modeling for social media analytics research is gaining traction (Jelodar et al., 2019) as it helps understand discussions and reactions of individuals participating in different social media sites. As a probabilistic topic modeling technique, the main idea behind the LDA is that "documents are represented as random mixtures over latent topics" and topics have a distribution over words (Jelodar et al., 2019). We specified six, ten, and fifteen topics in our LDA analysis and obtained similar results. Table 6 summarizes the results based on six topics with the terms associated with each topic generated by the LDA for the flu vaccine-related discussions and MMR vaccine-related discussions, respectively.

For the flu vaccine-related discussions, we interpreted the topics as the following based on the terms associated with each topic: (1) flu vaccine effectiveness and risks, (2) the timing of the flu vaccine, (3) flu vaccine strains and body immunity, (4) general discussions on flu vaccines, (5) flu vaccines and children, and (6) whether the flu vaccine is needed. Terms associated with the first topic show that the discussions focused on the effectiveness of the vaccine against flu viruses and their risks in causing infection. Topic 2 is concerned with need to get the flu vaccine every year. Topic 3 focused on the capability for the flu vaccine to protect against different strains of the flu viruses and cancer risk. Topic 4 included general discussions. Topic 5 focused on flu vaccines and children, and how being around children without the flu vaccine may get them sick. Topic 6 included discussions on whether there is a need for getting the flu vaccine.

For MMR vaccine-related discussions, we interpreted the discussions as the following based on the terms associated with each topic: (1) MMR vaccines beliefs, (2) autism risk linked to MMR vaccine on children, (3) timing of MMR vaccine, (4) research and evidence on MMR vaccines, (5) general discussions, and (6) effect of MMR vaccine on health. Topic 1 focused on the debate about MMR vaccine and some people who are against the vaccine. Topic 2 is concerned with the relationship between the MMR vaccine and autism. Some people argue that the vaccine could cause autism, while other people call for scientific evidence and research to prove this linkage. Topic 3 is related to the discussions on the timing of the MMR vaccine at one year of age. Topic 4 focused on research and scientific studies on MMR vaccines. Topic 5 included general discussions on the MMR vaccines. Topic 6 included discussions on the effectiveness of the MMR vaccine.

Table 6. LDA topic modeling results of all posts.

Topic	Terms									
	1	2	3	4	5	6	7	8	9	10
Flu vaccine-related discussions										
1. Flu vaccine effectiveness and risks	vaccin	effect	influenza	risk	season	also	disea	health	studi	caus
2. Timing of the flu vaccine	flu	shot	year	get	got	everi	week	last	month	even
3. Flu vaccine strains and immunity	can	immun	will	virus	still	strain	even	system	differ	bodi
4. General discussions	just	like	know	think	make	thing	realli	say	actual	right
5. Flu vaccines and children	get	people	sick	take	want	babi	die	don't	kid	around
6. Whether flu vaccine is needed	one	work	time	need	day	never	now	good	doctor	well
MMR vaccine-related discussions										
1. MMR vaccines beliefs	like	just	say	thing	think	even	actual	believ	reallo	way
2. Autism risk linked to MMR vaccine on children	vaccin	autism	studi	children	caus	risk	medic	mmr	link	parent
3. Timing of MMR vaccines	get	just	kid	know	time	doctor	got	said	shot	now
4. Research on MMR vaccines	one	use	also	mani	scienc	read	post	remov	never	tri
5. General discussions	peopl	can	will	make	need	want	person	reason	right	live
6. Effect of MMR vaccine on health	vaccine	measl	diseas	immun	case	mmr	effect	health	year	mump

We also performed topic modeling using latent semantic analysis (LSA). It is a computational model that builds semantic space from text corpus (Cvitanic, Lee, Song, Fu, & Rosen, 2016). It relies on the frequencies of co-occurrence of terms across different documents (Williams & Betak, 2018). This co-occurrence implies that these terms are related. In addition, the documents that include the related terms could be related to one group (Gefen, Endicott, Fresneda, Miller, & Larsen, 2017). After analyzing the pattern of words co-occurrence in many documents (samples), the semantic representations are created (Williams & Betak, 2018).

The LSA results for the flu vaccine based on cosine similarity revealed the following topics: (1) getting the flu vaccine every year, (2) flu vaccine effectiveness, (3) asking questions about the vaccine, (4) and general discussion about the flu vaccine. The results of the LSA analysis are consistent with the LDA analysis results. The results of the LSA topic modeling based on cosine similarity for MMR vaccine discussion revealed the following topics: (1) the need for studies and scientific research for the link between autism and MMR vaccine, (2) children getting two doses of the vaccine, (3) general discussion about the vaccine, and (4) the claimed negative consequences of the vaccine. As noticed, there is some overlap between the results of LDA and LSA.

2.6.4 Thread Initial Posts Analysis Results

In addition to analyzing all posts including the thread initial posts and replies discussed above, we performed text analytics for the initial post in each thread only. The results of the text analytics of the initial posts are discussed in the following section and are later used in our regression analysis to predict the number of replies for the thread.

2.6.4.1 Term Association Analysis Results of Thread Initial Posts. We used the same cutoff of 0.2 similar to our term association analysis of all posts. The terms associated with the flu vaccine in the initial posts are shown in Appendix A, and those associated with the MMR vaccine in the initial posts are shown in Appendix B. Compared with the term association analysis of all posts, more terms had a 0.2 or higher correlations in the analysis of the first post of each thread for both flu- and MMR-vaccine related discussions. These results suggest that discussions in the thread initiating posts were more focused using similar terms, while discussions in all posts including the replies were more diffused using different words. There were also more terms identified in the MMR vaccine initial posts than the flu vaccine initial posts, indicating discussions in the MMR vaccine initial posts were more focused.

2.6.4.2 Sentiment Analysis Results of Thread Initial Posts. Sentiment measures for the initial posts were extracted using LIWC and summarized in Table 7. The results of the sentiment analysis indicate that in the initial posts positive emotions were significantly different between the two vaccines' discussions. More positive emotions were expressed in the flu vaccine initial posts compared with those on the MMR vaccine. In addition, anger was also significantly different and expressed more in the flu vaccine initial posts.

Table 7. Sentiment analysis results for thread initial posts

Sentiment Measure	Flu Vaccine Initial Posts Mean	MMR Vaccine Initial Posts Mean	t-Stat on Difference in Means
Positive Emotions	2.065 (3.185)	1.211 (2.647)	3.199***
Negative Emotions	1.861 (3.162)	1.603 (2.346)	1.017
Anger	0.590 (2.005)	0.279 (1.133)	2.120**
Anxiety	0.579 (1.630)	0.520 (1.340)	0.429
Sadness	0.216 (1.292)	0.222 (1.006)	-0.051
Notes: Standard deviations in parentheses. *** p<0.01; ** p<0.05; * p<0.10.			

2.6.4.3 Topic Modeling Results of Thread Initial Posts. We also performed topic modeling for initial posts to examine the topics discussed by thread initiators. The results are reported in Table 8. For the flu vaccine-related initial posts, we interpreted the topics as the following based on the terms associated with each topic: (1) need for flu vaccine, (2) flu virus strains and flu vaccine protection for infants, (3) flu season and vaccine related studies, (4) need for annual flu vaccine, (5) flu vaccines and benefits for children, and (6) flu vaccine general discussion. For MMR vaccine-related initial posts, we interpreted the discussions as the following based on the terms associated with each topic: (1) studies about MMR vaccines, (2) MMR vaccine effects and risks on children, (3) MMR vaccine and autism, (4) general discussion about MMR vaccine, (5) MMR vaccine and children in Japan, and (6) MMR severe reactions in Japan.

Table 8. Topic modeling results for thread initial posts.

Topic	Terms									
	1	2	3	4	5	6	7	8	9	10
Flu vaccine-related discussions										
1. Need for flu vaccine	vaccin	one	babi	need	doctor	want	day	flu	get	people
2. Flu virus strains and flu vaccine protection for infants	flu	virus	protect	can	strain	still	cause	infant	prevent	die
3. Flu season and vaccine related studies	vaccin	find	season	new	studi	current	now	tumor	system	years
4. Need for annual flu vaccine	get	flu	year	people	shot	will	free	like	think	risk
5. Flu vaccine and benefits for children	vaccin	influenza	children	effect	report	also	develop	hospit	immun	include
6. Flu vaccine general discussion	got	just	shot	time	way	week	make	like	feel	said
MMR vaccine-related discussions										
1. Studies about MMR vaccines	vaccin	mump	measl	wakefield	medic	studi	article	lancet	rubella	disea
2. MMR vaccine effects and risks on children	mmr	vaccin	caus	link	immun	outbreak	risk	babi	month	sinc
3. MMR vaccine and autism	vaccine	autism	children	studi	associ	mmr	age	risk	measl	asd
4. General discussion about MMR vaccine	measl	get	just	rubella	like	can	know	doctor	will	make
5. MMR vaccine and children in Japan	vaccin	japan	health	children	infant	first	develop	two	countri	year
6. MMR severe reactions in Japan	vaccin	case	report	japan	adver	effect	reaction	one	state	death

2.7 Regression Analysis

To test Hypotheses 1-4, we examine the sentiment, topic, and use of high frequency words' impacts on the participation in a thread on Reddit using regression analysis. Participation is measured by the number of comments a first post received from users of Reddit. Because our dependent variable is a count variable and the variance (62,808.41) and the mean (60.79) of the comments count were very different, we used the negative binomial regression to control for over-dispersion (Chun, Leem, & Suh, 2021). Table 9 summarizes the variables definitions.

Table 9. Variables definitions

Variable	Definition
Comment_Count	The number of comments the thread initial post received.
Neg_Emo	The negative emotions score of the initial post of the thread.
Pos_Emo	The positive emotions score of the initial post of the thread.
MMR_Vaccine	The vaccine type of the thread; 0 for flu and 1 for MMR.
Weekend	A dummy variable to indicate whether the initial post of a thread was created on a weekday or a weekend; 0 for weekday and 1 for weekend.
Count_High_Corr_Words	The number of highly correlated words (identified for the vaccine through term association analysis in Appendices A1 and A2) used in the first post of the thread.
Initial_Post_Words_Count	The number of words in the initial post of the thread.
Subreddit	The subreddit that the thread was posted in.

We test two different models of comment count. Model 1 represents our negative binomial regression of the comment count with a dummy variable representing the vaccine type.

Model 2 further dissects the discussions based on the topic for each vaccine instead of using the vaccine type dummy variable. The two models are specified as follows:

$$Comment\_Count_i = \beta_0 + \beta_1 Neg\_Emo_i + \beta_2 Pos\_Emo_i + \beta_7 MMR\_Vaccine_i + \beta_4 Weekend_i + \beta_5 Count\_High\_Corr\_Words_i + \beta_6 \text{Initial\_Post\_Words\_Count}_i + \varepsilon_i \quad (1)$$

$$Comment\_Count_i = \beta_0 + \beta_1 Neg\_Emo_i + \beta_2 Pos\_Emo_i + \beta_3 Vaccine\_Topic_i + \beta_4 Weekend\_Dum_i + \beta_5 Count\_High\_Corr\_Words_i + \beta_6 \text{Initial\_Post\_Words\_Count}_i + \varepsilon_i \quad (2)$$

Table 10 shows the pairwise correlations of the variables.

Table 10. Pairwise correlations.

Variable	1	2	3	4	5	6	7
1. Comment_Count	1.00						
2. Neg_Emo	-0.04	1.00					
3. Pos_Emo	-0.05	-0.18	1.00				
4. Vaccine_type	0.04	-0.05	-0.14	1.00			
5. Weekend	-0.03	-0.04	-0.04	0.11	1.00		
6. Count_High_Corr_Words	-0.03	0.16	0.05	-0.37	-0.01	1.00	
7. Initial_Post_Words_Count	-0.08	0.01	-0.04	0.21	0.00	0.31	1.00

The results of Models 1 and 2 are shown in Table 11.

Table 11. Negative binomial regression results of comment count.

Variable	**Model 1**	**Model 2**
Intercept	5.298*** (0.297)	4.310*** (0.345)

Table 11, cont.

Neg_Emo	-0.085** (0.028)	-0.054 (0.028)
Pos_Emo	-0.070** (0.027)	-0.059* (0.026)
MMR_Vaccine	0.242 (0.164)	
Topic: Flu virus strains & flu vaccine protection for infants		0.586 (0.323)
Topic: Flu season and vaccine-related studies		0.354 (0.321)
Topic: Need for annual flu vaccine		-0.130 (0.361)
Topic: Flu vaccine and benefits for children		0.617 (0.345)
Topic: Flu vaccine general discussion		-0.603 (0.373)
Topic: Studies about MMR vaccines		-0.245 (0.384)
Topic: MMR vaccine effects and risks on children		0.922*** (0.275)
Topic: MMR vaccine and autism		0.833* (0.365)
Topic: General discussion about MMR vaccine		-0.404 (0.331)
Topic: MMR vaccine and children in Japan		0.406 (0.517)

Table 11, cont.

Topic: MMR severe reactions in Japan		0.168 (0.464)
Weekend	-0.421 (0.241)	-0.243 (0.240)
Count_High_Corr_Words	0.000 (0.001)	-0.000 (0.010)
Initial_Post_Words_Count	-0.324*** (0.081)	-0.128 (0.085)
N	474	474
AIC	4311.3	4303.1
Notes: Standard Error in parentheses. ***p < 0.001; **p < 0.01; *p < 0.05.		

The regression results indicate that emotions expressed in the discussion had negative impacts on users' participation in the thread. This result could be because when people express more emotions in their posts, they might not need answers from others and thus others might read their stories but prefer to reply to posts that need their opinions or answers. In addition, initial posts word count had a negative impact on the participation in the thread. Usually when people ask questions, they use fewer words and sentences to ask their questions compared with people telling their stories or experience or providing some information. Questions require others' participation and thus have more replies compared with other posts. In addition, social media content is very dense and crowded. People navigate through the posts and read and participate in some of them. Many people tend to select shorter posts with fewer words to engage in.

Regarding the topics of initial posts, some topics were shown to have more participation than others. For instance, the MMR vaccine effects and risks on children topic had a significant positive impact on participation in the thread. This indicates that people are interested in this topic and engaged in it. Another topic that impacted participation positively is the MMR vaccine and autism. The relationship between the MMR vaccine and autism is one of the topics that has people's attention online and offline. Individuals seek evidence and scientific proof of this information.

2.8 Discussion

2.8.1 Results Summary

We analyzed online discussions about flu and MMR vaccines on Reddit using different types of text analytics techniques. Term association analysis of Reddit threads reveals the terms that are highly associated with the flu vaccine including get, every, year, season, strain, and people. The terms that are highly associated with the MMR vaccine include vaccine, risk, autism, cause, and disease. The results imply that discussions about voluntary vaccines such as the flu vaccine is concerned with encouraging or discouraging people to take the vaccine since it is their choice to take the vaccine. However, the mandatory vaccine discussion is more focused on risks claimed to be related to the vaccination such as autism in the case of the MMR vaccine. In addition, since the MMR vaccine is given to children, the most discussed age group is children. The flu vaccine is given to different age groups so that the term people is associated with it in the online discussion.

Sentiment analysis shows that sentiment and affect expressed for the two vaccines are different. Flu vaccine discussions are shown to be more emotional both positively and negatively. Emotions such as anger, sadness, and anxiety are expressed in the discussion.

Topics discussed regarding the two vaccines varied as well. The effectiveness of the flu vaccine is one of the main topics that are common and diffused. The MMR vaccine risks and the need for evidence and credible information from popular health protection agencies such as the CDC are the most common topics discussed.

2.8.2 Theoretical Contribution

We examined how two different types of vaccines are discussed on online communities. Our research has the following contribution to theory. First, our study extends the online health communities' literature by examining their content which may reveal latent variables that could be difficult to identify through other means.

Second, our research is one of the first to compare and contrast online discussions of two types of vaccines: mandatory vs. voluntary using text mining including topic modeling, sentiment analysis and term association. Our results enable researchers to get a better insight about vaccine acceptance and rejection and the factors that impact them.

Third, our sentiment analysis results reveal the importance of emotions expressed in online discussion, which could also help in identifying anti-vaccine movement and improve the knowledge about the factors behind vaccine rejection such as individuals' personality treat, bad experience with vaccines, or misinformation. Misinterpretations of causality between MMR vaccine and autism is one example that many people debate. Evidence and research can help to negate these doubts and misinformation (Aghili & Lapointe, 2019).

Fourth, our research highlights how results from text mining can inform online community participation. Our prediction of thread participation using the sentiment, word count, and topic discussed in the initial post reveals the importance of these factors and the usefulness of text mining techniques in understanding social media user behavior.

2.8.3 Practical Implications

Our results highlight the importance of using online health communities and social networks to discuss various health-related issues such as vaccines. The online content helps health agencies to improve vaccine communication and addresses the concerns associated with vaccines to enhance vaccine confidence and eliminate vaccine hesitancy (Kang et al., 2017). Public healthcare strategies could leverage online content and big data power to be more effective and control the spread of infectious diseases and prevent dissemination of related misinformation (Brunson, 2013). Administrative agencies could employ suitable awareness campaigns for different types of vaccines to target the most concerned groups such as new parents for vaccines required for newborns and infants (Fukuda et al., 2014). Moreover, our content analysis reveals the importance of social media to disseminate required information from credible agencies so that misinformation could be prevented. The topic analysis shows the need for evidence of vaccine effectiveness and related risks. Health agencies could fulfill individuals' needs by providing them with this information and thus encourage them to take the vaccines.

2.9 Conclusion, Limitations, and Future Research Directions

Our study reveals that different types of vaccines have different online discussions. Mandatory vaccines have more online participation. Voluntary vaccines have more controversial discussion. Text mining of vaccine online discussion can help in understanding the concerns and beliefs related to vaccination. Our work has some limitations. First, we analyzed publicly available data using posts from Reddit. Other data such as the posts' structure and who replies to whom could be helpful to understand the nature of these social networks and their social influence. Future research could address this limitation. Second, our study focused on flu and MMR vaccines. These two vaccines are highly recommended to achieve personal and public protection, particularly for healthcare workers (Little et al., 2015). However, online conversations concerning other essential vaccines could be examined such as hepatitis B vaccine and COVID-19 vaccine. Finally, our study is limited to Reddit online communities. Future research could examine other online health communities and compare the contents from different platforms.

CHAPTER III

THE ROLE OF OPINION LEADERS IN ELEVATING ONLINE HEALTH DISCUSSION: EVIDENCE FROM REDDIT

3.1 Introduction

Online social networks and social networking sites (SNS) are stunning communication platforms (Bamakan, Nurgaliev, & Qu, 2019). Their use has grown exponentially (Hawi & Samaha, 2017). Social networking sites are used by billions of people all around the world (Hossain, 2019). Many individuals' lives depend on SNS and social media for communication with friends, expressing their opinion, looking for information, and sharing information and photos, and many others. One popular SNS is Reddit that is known by the "front page of the internet" (Baumgartner, Zannettou, Keegan, Squire, & Blackburn, 2020). It has distinguished characteristics such as anonymity (Rhidenour, Blackburn, Barrett, & Taylor, 2021) and topic-centered discussions (Donelson et al., 2021). Reddit ranks as the sixed most visited website in the United States (Eghtesadi & Florea, 2020). Users' participation on these SNS varies largely. Some users are active posters, others are active commenters, while others are limited to navigating available posts and comments and play as lurkers in the SNS. Opinion leaders play an effective role in disseminating information to others and affecting their behavior, attitudes, and opinions (Bamakan et al., 2019; Shi & Salmon, 2018).

Health-related discussion on social media is getting popular (Donelson et al., 2021; Lama, Hu, Jamison, Quinn, & Broniatowski, 2019). Many health problems are discussed in Reddit, including vaccines (Alazazi & Wang, 2021), mental health problems (Park et al., 2018), depression (Pirina & Çöltekin, 2018), chronic diseases (Park et al., 2018), and many others. Previous studies mainly focused on analyzing discussion topics and patterns (Fraga, da Silva, & Murai, 2018). Participants on these discussion forums seek health advice (Buntinx-Krieg, Caravaglio, Domozych, & Dellavalle, 2017), discuss the effectiveness of some medications and related side effects (Alazazi & Wang, 2021), provide support to community members (Fraga et al., 2018), and others.

Opinion leaders' role and interventions in health discussion are important to have interactive communication and affect the spread of health information (Yin, Xia, Song, Zhu, & Wu, 2020). The majority of previous studies that examined opinion leaders' impact on social media used a survey approach (Andrews, Tonkin, Lancastle, & Kirk, 2014; Bergström & Jervelycke Belfrage, 2018; Chu et al., 2019; Nisbet, 2006; Nunes, Ferreira, de Freitas, & Ramos, 2018; S. Y. Song, Cho, & Kim, 2017; Winter & Neubaum, 2016; Xiong, Cheng, Liang, & Wu, 2018). Furthermore, previous research using archival data has focused mainly on exploring opinion leaders on Twitter (Chu et al., 2019; Lamirán-Palomares, Baviera, & Baviera-Puig, 2019; Riquelme, Gonzalez-Cantergiani, Hans, Villarroel, & Munoz, 2019), Facebook (Oueslati et al., 2021; Winter & Neubaum, 2016), and other popular SNS such as the Chinese Sina-Weibo microblog (Yin et al., 2020)(Yang, Qiao, Liu, Ma, & Li, 2018). However, their role in SNS such as Reddit is not sufficiently studied. To fill this gap, this study aims to examine the different types of opinion leaders' impacts on the dynamics of online conversation including post volume.

3.2 Literature Review

3.2.1 Opinion leaders

The opinion leader concept is well-established in the literature (Bamakan et al., 2019). It is part of the two-step flow theory that proposes that information flows first from mass media and reaches opinion leaders (Lazarsfeld, Berelson, & Gaudet, 1944). Then these opinion leaders pass and transmit the information to the less active individuals in their network whom they influence (Lazarsfeld et al., 1944) (Elihu Katz, 1957). Lazarsfeld and his colleagues originally studied opinion leaders for the 1940 presidential elections, and found that opinion leaders influence decision making by others (Rogers & Cartano, 1962). Opinion leaders are influential and attractive users that disseminate information and diffuse ideas from media outlets to other individuals and impact their attitudes and perspectives and shape the public opinion later (E Katz & Lazarsfeld, 1955). Previous research emphasize the importance of social relationships in ideas' communication (Rogers & Cartano, 1962). Communication between individuals changes a society from several dispersed individuals to a body of interacting and connected people through personal influence and opinion leaders.

Opinion Leaders have distinguishable social and psychological aspects. They also have reliable knowledge in a specific field (Bamakan et al., 2019). Due to their knowledge and interpersonal and communication skills, opinion leaders become influential social participants. Opinion leaders are socially well-engaged and are considered credible and trustworthy by their followers (Xiong et al., 2018). They usually have a wide circle of acquaintances and social capital (Burt, 1999). Thus, they play the role of brokers who disseminate information and innovations between groups (Burt, 1999). Opinion leader are defined in the literature as "individuals who exert an unequal amount of influence on the decisions of others" (Flynn, Goldsmith, & Eastman, 1996). They are distinguished from their followers by having more accurate sources of information, global and more diverse social relationships, and more innovativeness (Rogers & Cartano, 1962).

Nowadays, with online social networks, opinion leaders can reach a larger number of people through social media outlets and disseminate information further than traditional social networks. They can build a tremendous social capital that could reach millions of followers (Winter & Neubaum, 2016).

3.2.2 Opinion leaders on social media

Opinion leaders use social media to disseminate information to others, through establishing social interactions with others in virtual communities (Xiong et al., 2018). Self-identity, knowledge contributions, and reciprocity are found to enhance opinion leaders' social interactions (Xiong et al., 2018). Other characteristics of opinion leaders on Facebook were identified such as personality strength and political interests (Winter & Neubaum, 2016). Another study examined personality traits of opinion leaders in social media (S. Y. Song et al., 2017). The study found that openness, exhibitionism (i.e., to be extraverted), and effective interactive ability were of the significant characteristics of opinion leadership. In addition, flow experience was found to strengthen and mediate the relationship between opinion leadership and opinion leadership behavior in social media.

The motives of opinion leaders on SNS are studied and include the interest in disseminating information, creating a positive impression to hold an influential social position, and persuading others (Winter & Neubaum, 2016). Furthermore, opinion leaders in online social networks play a vital role in e-commerce (Y. Zhao, Kou, Peng, & Chen, 2018) marketing

(Momtaz, Aghaie, & Alizadeh, 2011) and fashion industry (Casaló, Flavián, & Ibáñez-Sánchez, 2018). Word-of-mouth impacts opinions and consumer decision, particularly, when it is from influential users. For instance, opinion leaders were found to have a positive impact on buyers' intention to buy products that are evaluated by opinion leaders through persuasive messages (Nunes et al., 2018). Opinion leaders were able to change followers' attitude so that they accept the provided information about reviewed goods, which impacted their purchase intention. Opinion leaders' characteristics include reliability, competence, knowledge, empathy, previous experience, and others. Credibility of opinion leaders is vital to enhance the influence power of e-commerce (Y. Zhao et al., 2018). Uniqueness and originality are some aspects that significantly impact fashion opinion leadership on Instagram (Casaló et al., 2018). Opinion leadership has an impact on followers' intention to take influencer's advice.

Opinion leaders on social media are studied in many contexts. One study identified opinion leaders in online learning communities, where opinion leaders were active in discussions and had an obvious influence on other learners (Luo, Yang, Chen, & Wei, 2018). In addition, opinion leaders in Twitter were identified during sports events (Lamirán-Palomares et al., 2019). Some variables are used in their identification including indegree centrality and the number of followers. Indegree measure identified cyclists and sport institutions among the most influential users, while the number of followers identified general media and popular user accounts as the most influential users. The study did not find a single variable to be sufficient to identify opinion leaders in sports discussion implying the need for multi-variable identification. Furthermore, travel opinion leaders and seekers were identified and found to be highly connected to each other (Yoo, Gretzel, & Zach, 2011). Social media engagement, technology skills, education, young age group, and travel planning characterized both opinion leaders and seekers. However, opinion leaders are distinguished from opinion seekers by having greater travel experience that is more frequent and international and trust in official social media sources of tourism information.

Many studied attempted to identify opinion leaders on social media using automated algorithms and social network analysis (Yang Liu, Gu, Ko, & Liu, 2018; X. Song, Chi, Hino, & Tseng, 2007; Xiong et al., 2018; Zhai, Xu, & Jia, 2008). Improved weighted leader rank was proposed for online learning communities that adds user interactivity and initial influence (Luo et al., 2018). Another study proposed node importance analysis using importance matrix iterative method to detect opinion leaders in multi-relationship online social networks (Sun & Bin, 2018). In addition, a previous study identified opinion leaders in Twitter by proposing a new centrality measure called Milestone rank that accounts for two parameters: interest and exclusivity of users for a specific topic (Riquelme et al., 2019). Weighted milestones should be defined first for each topic to calculate the weighted rank. The approach could be used to identify opinion leaders in other SNS as well. Furthermore, the closeness algorithm is proposed to identify opinion leaders on SNS (Yang et al., 2018). The closeness method maps the relationship between the nodes based on their interaction types in the SNS. Non-adjacent nodes get a delay of information spread that is considered in calculating node closeness centrality.

3.2.3 Health discussion on social media

Online health communities and health discussion are having a wide participation by many individuals (K. Zhao et al., 2014). The reason behind this increased participation is the benefits gained including getting quick answers to health questions and getting and providing support and empathy to others, which in terms could reduce stress level and help individuals to be more optimistic (K. Zhao et al., 2014). Influential users on these communities have an obvious impact on their effectiveness through active participation and convincing messages dissemination.

Opinion leaders participation influences health attitudes and behavior through helping individuals in coping with their illnesses, encouraging healthy life style, adopting new treatment, and others (Mohamad et al., 2017; K. Zhao et al., 2014).

Opinion leaders in tobacco-related Twitter content are studied and found to use more tobacco products than their followers and general Twitter users (Chu et al., 2019). Moreover, HPV vaccine online content was evaluated to understand the knowledge structure of the young adults (Amith et al., 2020). Using distributional semantics and other social media analytics, the study found that young adults lack the important knowledge about HPV vaccine and its role in preventing ovarian cancer, and they were mostly concerned about the direct impact of the virus itself. Another study emphasized the role of the called “mommy bloggers” as opinion leaders on social networks to disseminate information to other mothers that promote awareness of HPV vaccination (Burke-Garcia, Berry, Kreps, & Wright, 2017).

In addition, opinion leaders’ role in the propagation of Covid-19 outbreak information and public health policies was examined in the Chinese Sina microblog (Yin et al., 2020). The study examined and compared temporal evolution of forwarding quantities between opinion leaders and other users. The results of the study emphasize the important role of opinion leaders in spreading information related to Covid-19 outbreak and influence the public opinion. In addition, online cancer support groups were analyzed (E. Kim et al., 2017). The results found that opinion leaders in these groups have active social interventions. They gained better psychosocial health outcomes including cancer information competence, better disease knowledge, and better coping strategies that reduces stress caused by their health condition and be more optimistic.

Personal and social attributes of opinion leaders in SNS Weibo related to organ donation were identified through the retweet network of organ donation messages (Shi & Salmon, 2018). Active users although unverified were found to accelerate the organ donation message sharing. In addition, experience and both medical knowledge and knowledge in Information technologies impacted user activeness and local opinion leadership about organ donation. The number of followers in the social network was also an important predictor of local opinion leadership. Another study analyzed social interactions in online health community about eating disorder and found that users tend to interact with others in the same community in Twitter rather than between communities (T. Wang, Brede, Ianni, & Mentzakis, 2018). The study compares two types of users: pro-recovery and pro-eater disorder. The two groups differ in their social behavior. Pro-eating disorder users have more negative emotions and feelings such as obvious feeling of social isolation and refusal indicating being at a higher risk of mental health problems. In addition, this group is found to have more active participation compared with pro-recovery group. The study finds that users with central position in the community can be considered opinion leaders and encourage others to adopt a specific lifestyle.

3.3 Hypothesis Development

In this study, we examine opinion leaders' participation in online health discussions and their impacts on post-interactivity and characteristics. Opinion leaders can be identified and classified into different categories according to their content, influence, and participation.

There is a need to explore the relationship between online public virtual spaces and the online interactions through them such as posting and message dynamics (Jones, Ravid, & Rafaeli, 2004). In this study, we aim to explore different types of opinion leaders' participation and their impact on the dynamics of threaded conversation including volume. Specifically, we examine how posts that engage opinion leaders receive attention from others. Conversation volume captures the size of the thread in terms of the number of posts in the thread. Previous research found that with interactive communication overload on social media, users tend to respond to simpler messages that they understand easily (Jones et al., 2004). Opinion leaders usually use simple language in their discussions and as a result, more people could engage in their discussion and generate posts and replies to that discussion. As a result, the conversation volume would increase. Opinion leaders write their posts in a creative clear way that gets the audience's attention. As a result, they respond to those posts and interact with them, either by commenting on them, liking them, or showing appreciation by providing rewards to the post such as the helpful reward and others.

In current contemporary SNS, a wide range of discussions and communication channels are available for users to participate in (Schäfer & Taddicken, 2015). These channels are used for informing people and communicating with friends and followers. Some research argues that in these circumstances, opinion leaders are vital (Schäfer & Taddicken, 2015). Since there are diverse social media outlets and discussions that are interconnected and linked, more orientation and guidance for the public is needed. However, opinion leaders have different communicative roles based on their characteristics, scope, number of issues, and social network structure (Bamakan et al., 2019). For health-related content, opinion leaders could persuade the audience about health-related habits or drugs (Oueslati et al., 2021). Opinion leaders usually have the skills and competency to create engaging comments. They are essential parties in the flow of communication on social networks (Lazarsfeld et al., 1944). The content created by them can encourage others to participate by discussing their point of view or adding vital aspects to the conversation. In addition, opinion leaders are connecting to a wider social network. Their comments or interventions in a thread help in passing the information to other individuals, who may not notice the post without the opinion leaders’ participation. Then, these individuals themselves could participate in the conversation and create their own content (Karlsen, 2015). Opinion leaders are classified into local and global ones based on their participation and influence (Bamakan et al., 2019). Local opinion leaders focus their engagement inside a specific online community or discussion group. Global opinion leaders engage in multiple diverse communities building a larger follower network. They have distributed participation in different discussion groups. Both types of leadership have an impact on information flow in social media, however, they differ from each other (Shi & Salmon, 2018). Local opinion leaders have a direct influence on one’s neighbors, while global opinion leaders have an indirect influence that allows for broader information exchange. Local and global opinion leaders influence their environment and information exchange. However, each of them has a different scope of influence (Oueslati et al., 2021). Global opinion leaders could have a more general perspective and consider the global community when discussing health-related content (Q. Xu, Yu, & Song, 2018). While local opinion leaders could have a narrower point of view depending on the local issue focusing their discussion on certain aspects. As a result, each of them will have a different impact on the discussion thread dynamics. Thus, we hypothesize:

H1a: Local opinion leaders’ participation in a discussion thread is positively associated with user thread participation.

H1b: Global opinion leaders’ participation in a discussion thread is positively associated with user thread participation.

Social media these days are crowded with large volumes of content makes it impossible for individuals to scan and participate in all of them (Karlsen, 2015). As a result, they filter the content using different techniques. Content created by opinion leaders gets an additional attraction by individuals to view and participate because of the characteristics of opinion leaders that make their content more visible. Based on the two-step flow of communication identified by Lazarsfeld et al. (1944), opinion leaders get the information from media outlets and then disseminate it to their acquaintances. As a result, opinion leaders serve as a source of new information and their content would be trendy and attractive to others (Winter & Neubaum, 2016). They make the information more accessible to ordinary users. In addition, opinion leaders select what to discuss in a post and what is considered an important issue to the reader (Q. Xu et al., 2018). Health-related content is continuously changing. There are always new research studies, diseases, medications, and recommended health practices. Thus, opinion leaders create health content on social media with interesting new information that would attract their friends and followers and encourage more interaction with it. Moreover, due to the prominent position of both local and global opinion leaders in the social network, their content would propagate to a wide range of people who are connected to them directly and indirectly (Hou, 2022). Thus, we hypothesize:

H2a: A local opinion leader-initiated thread is positively associated with user thread participation.

H2b: A global opinion leader-initiated thread is positively associated with user thread participation.

According to the content of opinion leader participation, there are both positive constructive and negative destructive opinion leaders. their contents is either positively or negatively emotionally charged. Studies find that emotional messages are more disseminated on social media outlets than neutral ones (Stieglitz & Dang-Xuan, 2013). Positive constructive opinion leaders are users who use persuasiveness, good language, and commitment to express their content (Bamakan et al., 2019). Constructive opinion leaders support their followers and help them cope with their struggles and health problems. Furthermore, when talking about a health-related issue, users could support or oppose it. Opinion leaders' attitudes and sentiment could be positive or negative (Q. Xu et al., 2018). In health-related discussions, people seek emotional support and optimistic thoughts to relieve their stress and concerns and find solutions (Franke, Felfe, & Pundt, 2014). As a result, they would interact with posts that have a positive impact. Positive opinion leaders foster a supportive climate and encourage a healthy lifestyle. Thus, we hypothesize:

H3a: A thread with constructive comments from local opinion leaders is positively associated with user thread participation.

H3b: A thread with constructive comments from global opinion leaders is positively associated with user thread participation.

On the other hand, destructive opinion leaders use language that has negative sentiments or attitudes. This type of opinion leader has influence; they utilize social media to disseminate their negative thoughts or sentiment about the subject discussed. They take advantage of their strong personality and communication skills to influence others' opinions. One study about educational Common Core state standards tweets found that the majority of opinion leaders expressed negative sentiment related to the Common Core (Y. Wang & Fikis, 2019). Negative sentiment posts were found to induce more comments compared to positive posts on social media outlets such as Facebook (Mayshak, Sharman, Zinkiewicz, & Hayley, 2017; Stieglitz &

Dang-Xuan, 2013). When a post has affect elements, particularly negative sentiment, it attracts the reader and requires his/her cognitive involvement and thus arouses his involvement with the post. For health-related online discussions, many people express negative feelings or concerns about the health problem and the risks associated with it. For instance, during the Covid-19 pandemic, many social media users spread their emotions and expressed their opinions. Having a new virus with uncertain aspects makes it easier to disseminate rumors and negative feelings (X. Xu, Li, Wang, & Zhao, 2021). Opinion leaders have communication competency; thus they are vital players in disseminating information and sentiment. Brand companies employ opinion leaders on social media to foster public engagement and create electronic word-of-mouth (Zhou, Barnes, McCormick, & Cano, 2021). Both local and global opinion leaders have their central positions in the social network and are connected to many other users. When using negative sentiment in their messages, opinion leaders affect others' feelings and trigger responses by them, either by supporting or opposing their emotional content. In addition, online health discussions foster a supportive culture (Carron-Arthur, Ali, Cunningham, & Griffiths, 2015). When focal people such as opinion leaders use negative sentiment to express their health problems, others will be stimulated to participate as a way of expressing their support or loyalty to these influential users. For example, using depressed language with negative feelings towards an illness such as the Covid-19 virus influences the emotional sentiment of others (K. Zhao et al., 2014). This encourages user intervention, which causes active participation in the thread. This later participation could have positive or negative emotions. Thus, we hypothesize:

H4a: A thread with destructive comments from local opinion leaders is positively associated with user thread participation.

H4b: A thread with destructive comments from global opinion leaders is positively associated with user thread participation.

3.4 Data and Methodology

To test our hypotheses, we collected online health discussion data on the popular social network Reddit. Reddit was founded in 2005 and amassed 355 million monthly worldwide active users in 2018 (Puri et al., 2020). In 2019, 11% of U.S. adults used the platform (Puri et al., 2020). According to Pew Research Center, social media use statistics showed that Reddit was one of the two platforms that had a significant user growth from 11% in 2019 to 18% in 2021, while most other popular platforms showed a slight growth (Auxier & Anderson, 2021). In 2021, Reddit ranked among the ten most popular social networking sites in the United States (Auxier & Anderson, 2021). The number of Reddit users has surged. In 2021, Reddit reported more than 50 million active users daily on its website.

Reddit is a unique social networking site that combines online communities called subreddits. Reddit users create different subreddits to discuss topics and issues of interest (Pirina & Çöltekin, 2018). In contrast to many SNS, users on Reddit are anonymous, which encourages users to speak more frankly and express their opinion more freely. Users create posts on Reddit that are upvoted or downvoted by other users (Guimaraes, Balalau, Terolli, & Weikum, 2019). The total votes of a post after a short time of its creation determines the visibility of the post and where it will show on the page (Haralabopoulos, Anagnostopoulos, & Zeadally, 2015). Another unique characteristic of Reddit is "karma", which is a user score that is based on his/her posting, commenting, awardees received, and awards provided activities.

Data has been scrapped from the "Mental Health" subreddit about different conversation threads related to Covid-mental health problems. Historical data in the period from January to August 2021 were scraped in which initial posts with the keywords "COVID", "coronavirus", or "corona virus" were collected. We also collected thread-related data such as initial post timestamp, author, post text, replies, and replies' authors. The initial dataset included 11,999 posts. Posts with deleted authors or missing textual content were eliminated. In addition, auto moderator posts that were generated automatically by the Reddit platform were eliminated. The final dataset after preprocessing and cleaning included 5,660 posts.

Social network analysis was performed to identify opinion leaders and their characteristics. A social network is defined as "*a set of actors, other entities, and a set or sets relations defined on them*" (Knoke & Yang, 2019). Network structure affects people's behavior and psychology. Social networks help us to better understand people's behavior and decisions (Knoke & Yang, 2019). In addition, social networks could be the causes and consequences of individuals' actions. Thus, social network analysis is a vital method that has an increased application in social sciences (Serrat, 2017). It could be applied to various fields such as business, health, online communications, and others. It focuses on understanding actors (nodes) and the structure of relationships between them in a particular context. Recently, leadership network analysis is gaining traction in understanding leaders' relationships within and across organizations or other groups (Serrat, 2017).

Social networks of posts and comments of different users were analyzed. Centrality measures were calculated to identify opinion leaders in the network and study their impact. Local leadership was measured by the degree centrality of a user in the network. Global leadership was measured by the eigenvector centrality of a user in the network. Constructive and destructive opinion leadership were measured through text mining of the content generated by these users. The Linguistic Inquiry and Word Count (LIWC) software was used to calculate positive and negative emotions and phrases in opinion leaders' posts.

Econometric analysis and machine learning algorithms were applied to test the proposed hypotheses. The effects of opinion leaders' engagement in the conversation at time t-1 on the change of thread conversation volume at time t were tested. The thread participation was measured as the count of new posts at time t that are added to the thread since time t-1. Control variables such as the subreddit the post was posted in, topic, year, and month of the year were also included.

3.4.1 Variables and Measures

Both local and global opinion leaders who participated in each thread were identified using social network analysis and centrality measures. Local opinion leaders were measured by the indegree centrality. Degree centrality is a measure of local centrality since it is measured by calculating the links between a focal node and its direct neighbors in the network without examining the global shape of the graph (Bamakan et al., 2019; Lamirán-Palomares et al., 2019). It indicates the number of people a person (node) has directly communicated with and thus has an influence on them (Carron-Arthur et al., 2015). In-degree centrality measures the number of incoming links to the node. It is an indicator of local opinion leadership (Bamakan et al., 2019). On the other hand, eigenvector centrality considers a node to be central when it is connected to other important nodes in the network. It assigns scores to all nodes in the network. And the centrality of a focal node is affected by the scores of the other nodes that it is connected to (Lamirán-Palomares et al., 2019). Thus, it is used as a measure for global opinion leaders in the conversation network.

R programming language was used to create the social networks and calculate the centralities using the “igraph” package. A directed network was created for the previous period of the current studied month. The opinion leaders in each month were identified based on the social network graph in the previous months. The nodes with the highest 10% indegree and eigenvector centralities were identified as local and global opinion leaders, respectively. In addition, threads created by opinion leaders were identified.

To identify constructive and destructive opinion leaders, sentiment analysis was performed. The LIWC application was used to study emotional aspects of posts created by local and global opinion leaders. LIWC has an internal dictionary that gives scores for text based on different word use (James W. Pennebaker & Kayla Jordan, 2015). Words contained in the text file were read and processed, then they were given scores based on the match with the internal dictionary words and categories. Both positive and negative sentiments of the texts of all posts were calculated. Our dependent variable, thread participation, was measured by the count of the comments in a thread between time t-1 and time t. It is a count variable with non-negative values.

3.4.2 Models Specification and Estimation

We constructed a short and unbalanced panel dataset of different threads spanning six-hour periods to examine how opinion leaders’ activities during a previous six-hour window affect the number of new replies to the thread during the next six-hour window. In our study, the dependent variable had a mean of 0.8258 and a variance of 3.0256, so we used the negative binomial regression model to resolve the over-dispersion issue of the dependent variable (Winkelmann, 2000). In addition, since the dependent variable was a count with excess zero values (59% of the values), we used the zero-inflated negative binomial (ZINB) regression

(Moghimbeigi, Eshraghian, Mohammad, & Mcardle, 2008). The ZINB is a generalized linear model that consists of two distributions, one for the zero values and the other for the non-zero values (Minami, Lennert-Cody, Gao, & Román-Verdesoto, 2007). It has a logit and a negative binomial regression.

For the negative binomial regression, we used a panel dataset that included the following independent variables: “local_ol_author”, “global_ol_author”, “Lag1_last_local_ol_count”, “Lag1_last_global_ol_count”, “Lag1_avg_posemo_local_opl”, “Lag1_avg_negemo_local_opl”, “Lag1_avg_posemo_global_opl”, and “Lag1_avg_negemo_global_opl”. We used lagged variables from the previous six hours period (t-1) to account for the endogeneity issue. We controlled for factors that could be important for explaining the health thread participation including “first_post_posemo”, “first_post_negemo”, “total_6hours_difference”, “month”, “time_of_day”, and “weekend”. The description of the variables is shown in Table 12.

Table 12. Regression variables description

Variable	**Description**
Dependent Variable	
Thread Participation	The number of new comments in a thread i between time t-1 and time t.
Independent Variables	
Local_ol_author	1 if thread i was initiated by a local opinion leader, 0 otherwise.
Global_ol_author	1 if thread i was initiated by a global opinion leader, 0 otherwise.
Lag1_last_local_ol_count	The number of comments added to thread i by local opinion leaders during time t-1.
Lag1_last_global_ol_count	The number of comments added to thread i by global opinion leaders during time t-1.
Lag1_avg_posemo_local_opl	The average positive emotion of local opinion leaders’ comments added to thread i during time t-1.
Lag1_avg_negemo_local_opl	The average negative emotion of local opinion leaders’ comments added to thread i during time t-1.
Lag1_avg_posemo_global_opl	The average positive emotion of global opinion leaders’ comments added to thread i during time t-1.

Table 12, cont.

Lag1_avg_negemo_global_opl	The average negative emotion of global opinion leaders' comments added to thread i during time t-1.
Control Variables	
First_post_posemo	The positive sentiment value of the first (initial) post of thread i.
First_post_negemo	The negative sentiment value of the first (initial) post of thread i.
Total_6hours_difference	The number of six-hour time periods that have elapsed between the first post in thread i and time t-1.
Month	The month that time t was in.
Time_of_day	The time period that time t was in; four values from 12 am to 6 am, 6:01 am to 12 pm, 12:01 pm to 6 pm, and 6:01 pm to 11:49 pm.
Weekend	1 if thread i was initiated on a weekend, 0 otherwise.

We used Stata to perform the regression analyses. All pairwise correlations were below 0.8, and all variance inflation factors (VIFs) were below 5, indicating no severe multicollinearity. Four models specified in Equations 1 through 4 below were examined. Model 1 included the control variables only. Model 2 included the control variables and the local opinion leaders' variables. Model 3 included the control variables and the global opinion leaders' variables. Model 4 included both local and global opinion leaders' variables in addition to the control variables.

$$Thread\ Participation_{it} = \beta_0 + \beta_1 FirstPostPosEmo_i + \beta_2 FirstPostNegEmo_i + \beta_3 Total6hoursDifference_{it-1} + \beta_4 Month_i + \beta_5 TimeOfDay_{it-1} + \beta_6 Weekend_i + \alpha_i + \varepsilon_{it} \qquad (1)$$

$$Thread\ Participation_{it} = \beta_0 + \beta_1 FirstPostPosEmo_i + \beta_2 FirstPostNegEmo_i + \beta_3 Total6hoursDifference_{it-1} + \beta_4 Month_i + \beta_5 TimeOfDay_{it-1} + \beta_6 Weekend_{it-1} + \beta_7 LastLocalOlCount_{it-1} + \beta_8 AvgPosEmoLocalOpl_{it-1} + \beta_9 AvgNegEmoLocalOpl_{it-1} + \beta_{10} LocalOlAuthor_i + \alpha\_i + \varepsilon\_it \qquad (2)$$

$$Thread\ Participation_{it} = \beta_0 + \beta_1 FirstPostPosEmo_i + \beta_2 FirstPostNegEmo_i + \beta_3 Total6hoursDifference_{it-1} + \beta_4 Month_i + \beta_5 TimeOfDay_{it-1} + \beta_6 Weekend_i + \beta_7 LastGlobalOlCount_{it-1} + \beta_8 AvgPosEmoGlobalOpl_{it-1} + \beta_9 AvgNegEmoGlobalOpl_{it-1} + \beta_{10} GlobalOlAuthor_i + \alpha\_i + \varepsilon\_it \qquad (3)$$

$$Thread\ Participation_{it} = \beta_0 + \beta_1 FirstPostPosEmo_i + \beta_2 FirstPostNegEmo_i + \beta_3 Total6hoursDifference_{it-1} + \beta_4 Month_i + \beta_5 TimeOfDay_{it-1} + \beta_6 Weekend_{it-1} + \beta_7 LastLocalOlCount_{it-1} + \beta_8 AvgPosEmoLocalOpl_{it-1} + \beta_9 AvgNegEmoLocalOpl_{it-1} + \beta_{10} LocalOlAuthor_i + +\beta_{11} LastGlobalOlCount_{it-1} + \beta_{12} AvgPosEmoGlobalOpl_{it-1} + \beta_{13} AvgNegEmoGlobalOpl_{it-1} + \beta_{14} GlobalOlAuthor_i + \alpha_i + \varepsilon_{it} \qquad (4)$$

α_i captures the fixed effect of thread i, β denotes model parameters to estimate, and ε_{it} is the error term.

3.5 Discussion

3.5.1 Results Summary

Table 13 summarizes our models' testing results for both the negative binomial regression and the logit model for predicting the probability that the thread had zero participation. Model 1 shows coefficient estimates for the control variables on thread participation. Models 2 and 3 show the coefficient estimates for local opinion leader- and global opinion leader-related variables, respectively. Model 4 has both types of opinion leader-related variables. The time elapsed since the thread initiation was negative and statistically significant in all four models for both the negative binomial and the zero inflated parts, except for the zero inflated part in Model 3. In addition, both positive and negative emotions expressed in the first post of a thread were found to have significant and positive associations with thread participation in the zero inflation results in Models 1 and 3. The number of local opinion leaders who participated in the thread in the previous period had a positive relationship with thread participation in the current period in Model 2 only. Thus, H1a was partially supported. Global opinion leaders' participation in a thread positively impacted new thread participation in Models 3 and 4, which supports H1b. Threads initiated by local opinion leaders were negatively associated with thread participation in Models 2 and 4. As a result, H2a was not supported. The results indicate that local opinion leader-initiated threads had less participation. This result indicates that local opinion leaders could be better commenters than thread initiators. In addition, global opinion leaders'-initiated threads were not significant except for Model 3's zero inflation part, thus H2b was not supported.

Emotions by opinion leaders had an impact on the discussion. Negative emotions and sentiments expressed by global opinion leaders had a positive and significant association with thread participation by Reddit members. Thus, H4b was supported, while H4a was not. The results show that people tend to participate more in discussion or threads that has negative emotions by global opinion leaders than positive ones. Local opinion leaders' expressed emotions did not impact following participation in the thread. As a result, H3 was not supported.

Table 13. Zero-Inflated negative binomial panel estimates for thread participation models

Variable	Model 1		Model 2		Model 3		Model 4	
Negative Binomial Regression								
	Coefficient (Std. Err.)	Incidence Ratio	Coefficient (Std. Err.)	Incidence Ratio	Coefficient (Std. Err.)	Incidence Ratio	Coefficient (Std. Err.)	Incidence Ratio
First_post_posemo	0.020 (0.031)	1.020	-0.003 (0.033)	.997	0.026 (0.032)	1.026	0.027 (0.032)	1.027
First_post_negemo	-0.002 (0.024)	.998	-0.020 (0.027)	.980	0.021 (0.024)	1.021	-0.003 (0.028)	.997
Total_6hours_difference	-0.016*** (0.002)	.984***	-0.010*** (0.001)	.990***	-0.015*** (0.002)	.986***	-0.009*** (0.001)	.991***
Local_ol_author			-1.413*** (0.194)	.243***			-1.299*** (0.200)	.273***
Lag1_last_local_ol_count			0.458** (0.204)	1.581**			-0.005 (0.231)	.995
Lag1_avg_posemo_local_opl			-0.005 (0.039)	.995			0.035 (0.042)	1.035
Lag1_avg_negemo_local_opl			0.039 (0.034)	1.039			-0.036 (0.042)	.964
Global_ol_author					-0.415 (0.673)	.660	-0.552 (0.459)	.576
Lag1_last_global_ol_count					0.651*** (0.121)	1.917***	0.645*** (0.127)	1.907***
Lag1_avg_posemo_global_opl					-0.038 (0.039)	.962	-0.054 (0.041)	.948
Lag1_avg_negemo_global_op l					0.059** (0.029)	1.061**	0.077** (0.032)	1.080**
Constant	0.065 (0.157)	1.067	0.364** (0.169)	1.439**	0.058 (0.163)	1.059	0.346** (0.169)	1.414**
Month	Included	Included	Included	Included	Included	Included	Included	Included
Time_of_day	Included	Included	Included	Included	Included	Included	Included	Included
Weekend	Included	Included	Included	Included	Included	Included	Included	Included
Zero Inflation								

Table 13, cont.

	Coefficient (Standard Error)	Odds Ratio	Coefficient (Standard Error)	Odds Ratio	Coefficient (Standard Error)	Odds Ratio	Coefficient (Standard Error)	Odds Ratio
First_post_posemo	1.043*** (0.289)	2.839***	0.051 (0.095)	1.052	0.515** (0.222)	1.674**	0.134 (0.094)	1.143
First_post_negemo	0.322** (0.156)	1.380**	-0.047 (0.081)	0.954	0.469*** (0.147)	1.598***	-0.0209 (0.088)	0.979
Total_6hours_difference	0.011** (0.005)	1.011**	0.026*** (0.004)	1.027***	0.002 (0.004)	1.002	0.024*** (0.005)	1.024***
Local_ol_author			-29.460 (871.500)	1.61e-13			-51.280 (1.076e+08)	5.34e-23
Global_ol_author					3.530** (1.677)	34.12**	0.764 (1.182)	2.147
Constant	-5.773*** (1.581)		-0.993** (0.440)		-4.283*** (1.216)		-1.022** (0.437)	
Month	Included	Included	Included	Included	Included	Included	Included	Included
Time_of_day	Included	Included	Included	Included	Included	Included	Included	Included
Weekend	Included	Included	Included	Included	Included	Included	Included	Included
N	3,037		3,037		3,037		3,037	
Log-likelihood	-2246.573		-2213.965		-2204.331		-2186.235	
χ^2	162.770***		199.910***		230.820***		247.030***	
*** p<0.01, ** p<0.05, * p<0.1								

3.5.2 Machine Learning

Machine learning algorithms are powerful tools to predict patterns in big data (Alazazi, Wang, & Allan, 2020; El Naqa & Murphy, 2015). Machine learning is a popular branch of artificial intelligence that uses mathematical tools to predict an outcome. Supervised machine learning is one type of machine learning that requires a labeled training dataset that includes both input and output variables (Alazazi et al., 2020). Different machine learning algorithms are used to uncover relationships within data. The algorithms include linear regression, logistic regression, support vector machine, decision tree, random forest, neural networks, and others. The performance of these algorithms varies depending on their accuracy of predicting the outcome variable.

In this study we employed some effective and powerful supervised machine learning algorithms to predict the thread participation in the online health discussion. We examined all previous four models using different algorithms to find out which combination of model and algorithm has the best prediction. We briefly summarize the machine learning algorithms used in the following section:

Poisson Regression: It is an appropriate statistical regression to predict count outcome variables. Our dependent variable is the count of comments during the current time period, which is a count variable. Thus, we use Poisson regression as the base statistical model to compare the other models with.

K-Nearest Neighbors: It is a nonparametric supervised algorithm that categorizes the training dataset cases into different classes so that the unlabeled data classes can be predicted. The class of the predicted data is selected based on some characteristics and the classes its closet neighbors from the training data set belong to. The K closest neighbors are determined to select the class for the new data point (Taunk, De, Verma, & Swetapadma, 2019).

Support Vector Machine (SVM): it is a hyperplane-based classifier built on the structural risk minimization principle of the statistical learning theory. It uses a small, limited number of samples from the training dataset to find the best classification, which allows for minimal sample point error. It selects the smallest classification surface as the optimal solution by finding the optimal hyperplane of the dataset that classifies the data points efficiently and has maximum spacing between two classes of data (Y. Zhang, 2012).

Neural Networks: The artificial neural networks mimic the human brain and can adapt to changing input. It consists of interconnected nodes (neurons) within multiple layers: input, hidden, and output. The hidden layers include combinations of some or all the predictor variables (Kuhn & Johnson, 2013). Its structure allows for continuous improvement by self-learning from previous mistakes. It uses feedforward mechanism where the output of one layer is the input data of the next layer in the network. In addition, it assigns weights to the different variables to determine their importance for the prediction. In this algorithm, the output variable is known and is compared with the predicted output variable by neural networks to selected the best parameters with the least error (Mahesh, 2020).

Random Forest: It is an ensemble data model that consists of multiple decision trees. Each tree is based on a random sample of a training data and a random subset of variables (Yanli Liu, Wang, & Zhang, 2012). It combines the results of different decision trees to get the most accurate classification or prediction outcome (Alazazi et al., 2020). It has the ability to measure the importance of each variable for prediction through model training.

R program was used to run different machine learning algorithms for all four models. The performance of each model was evaluated. Since we are predicting numeric values, root mean squared error (RMSE), R squared (coefficient of determination), and adjusted R squared metrics were reported for each algorithm and model to measure the performance during training and testing. Root mean squared error is “a function of the model residuals”. It measures the difference between the true values and the predicted values (Kuhn & Johnson, 2013). R squared measures “the proportion of the information in the data that is explained by the model” (Kuhn & Johnson, 2013). In other words, it measures how much of the variation of the outcome variable is explained by the model. Adjusted R squared considers the number of predictors in the model to overcome overfitting of the data problem. The results are reported in Table 14.

Table 14. Machine learning algorithms results

	Model 1 Controls			Model 2 Local OP.L.			Model 3 Global OP.L.			Model 4 All		
	RMSE	R^2	Adjusted R^2	RMSE	R^2	Adjusted R^2	RMSE	R^2	Adjusted R^2	RMSE	R^2	Adjusted R^2
Poisson	1.406	0.105	0.096	1.530	0.084	0.067	1.381	0.163	0.147	0.958	0.235	0.216
NNET	1.308	0.225	0.218	1.294	0.207	0.192	1.286	0.275	0.262	0.910	0.309	0.293
SVM	1.459	0.036	0.026	1.381	0.096	0.079	1.387	0.156	0.140	0.992	0.179	0.159
KNN	1.171	0.379	0.373	1.225	0.289	0.276	1.265	0.298	0.285	0.913	0.304	0.288
Random Forest	0.822	0.694	0.691	0.972	0.553	0.544	1.040	0.525	0.516	0.801	0.464	0.451

For all four models, random forest algorithm had the least RMSE, and the highest R-squared and adjusted R-squared values compared with other algorithms. Based on the evaluation metrics, random forest performed the best between all six algorithms to predict the number of comments on a thread at time t.

3.5.3 Robustness Checks

Additional tests and alternative measures were conducted to check the robustness of the results. First, in our main analysis, we chose the six-hours' time window for our panel dataset. We selected different time periods for the panel dataset including four-hours and twelve hours periods to re-estimate the models. The estimations confirm the main results of the study. For the 12-hours window, the local opinion leaders thread initiators variable and the negative emotions by global opinion leaders were marginally significant. However, the global opinion leaders' comments in the previous period were not significant. Table 15 summarizes the results of different time windows.

Second, in our main analysis, we selected last period variables for the count of the local and global opinion leaders and the positive and negative emotions expressed by them. The variables counted the comments in the previous t only without accumulating them. To check the robustness of our models, we used accumulated variables for the comments counts of both local and global opinion leaders in addition to the positive and negative emotions expressed by them. The results are shown in table 16. The results are consistent, and the coefficients estimates retained the same direction and significance as the main models of the study, except for the negative emotions by global opinion leaders. They were not significant, however, for the negative emotions by local opinion leaders they were significant in the cumulative models.

Third, local and global opinion leaders were identified using social network analysis as the nodes with the highest 10% of indegree and eigen vector centralities, respectively. We used different percentages of the nodes to identify the opinion leaders. We tried the highest 1%, 5%, and 20% of the nodes. Mainly, the results are consistent with the 10% opinion leaders. More details are shown in Appendix C.

Table 15. Results for models with different time windows

Variable	Model 4, 12_hours	Model 4, 6_hours	Model 4, 4_hours
Negative Binomial Part			
First_post_posemo	0.0174 (0.0694)	0.027 (0.032)	0.0301 (0.0292)
First_post_negemo	0.0320 (0.0584)	-0.003 (0.028)	-0.0211 (0.0252)
Total_hours_difference	-0.00776*** (0.00194)	-0.009*** (0.001)	-0.00580*** (0.000704)
Local_ol_author	-0.610* (0.319)	-1.299*** (0.200)	-1.307*** (0.177)
Lag1_last_local_ol_count	-1.095* (0.587)	-0.005 (0.231)	0.426* (0.231)
Lag1_avg_posemo_local_opl	0.0191 (0.0471)	0.035 (0.042)	-0.00900 (0.0540)
Lag1_avg_negemo_local_opl	-0.0364 (0.0771)	-0.036 (0.042)	-0.0232 (0.0274)
Global_ol_author	-0.415 (0.543)	-0.552 (0.459)	-0.689* (0.394)
Lag1_last_global_ol_count	0.393 (0.329)	0.645*** (0.127)	0.631*** (0.141)
Lag1_avg_posemo_global_opl	0.00230 (0.0533)	-0.054 (0.041)	-0.0198 (0.0441)
Lag1_avg_negemo_global_opl	0.102* (0.0566)	0.077** (0.032)	0.0504 (0.0325)
Constant	-0.399 (0.406)	0.346** (0.169)	0.0298 (0.172)
Month	Included	Included	Included
Time_of_day	Included	Included	Included
Weekend	Included	Included	Included
Zero Inflated Part			
First_post_posemo	0.302* (0.171)	0.134 (0.094)	0.0316 (0.0765)
First_post_negemo	-0.0494 (0.153)	-0.0209 (0.088)	-0.0248 (0.0705)

Table 15, cont.

Total_hours_difference	-0.00560 (0.00965)	0.024*** (0.005)	0.0199*** (0.00289)
Local_ol_author	-12.16 (557.9)	-51.280 (0.000)	-32.82 (1,051)
Global_ol_author	2.184* (1.261)	0.764 (1.182)	0.354 (0.941)
Constant	-0.788 (0.885)	-1.022** (0.437)	-0.366 (0.387)
Month	Included	Included	Included
Time_of_day	Included	Included	Included
Weekend	Included	Included	Included
N	1,043	3,037	4,926
Log-likelihood	-599.191	-2186.235	-3078.649
χ^2	55.09***	247.030***	303.47***

Table 16. Results with cumulative variables models

Variables	**Model 1**	**Model 2**	**Model 3**	**Model 4**
Negative Binomial				
first_post_posemo	0.020	0.008	0.014	0.034
	(0.032)	(0.034)	(0.033)	(0.033)
first_post_negemo	-0.002	-0.029	0.019	0.032
	(0.024)	(0.027)	(0.025)	(0.024)
total_6hours_difference	-0.016***	-0.010***	-0.015***	-0.013***
	(0.002)	(0.001)	(0.002)	(0.002)
1.local_ol_author		-1.334***		-0.969***
		(0.228)		(0.225)
Lag1_Cumulative_Local_PosEmo		0.040		0.052
		(0.032)		(0.035)
Lag1_Cumulative_Local_NegEmo		0.054**		0.064*
		(0.023)		(0.034)
Lag1_total_local_ol_count		-0.083		-0.408***
		(0.119)		(0.132)
1.global_ol_author			-0.446	0.140
			(0.655)	(0.644)
Lag1_Cumulative_Global_PosEmo			0.020	0.006
			(0.036)	(0.037)
Lag1_Cumulative_Global_NegEmo			0.012	-0.020
			(0.023)	(0.033)
Lag1_total_global_ol_count			0.170***	0.249***
			(0.058)	(0.065)
Constant	0.065	0.363**	0.047	0.111
	(0.157)	(0.173)	(0.168)	(0.171)
Month	Included	Included	Included	Included
Time_of_day	Included	Included	Included	Included
Weekend	Included	Included	Included	Included
zero-inflated				

Table 16, cont.

first_post_posemo	1.043***	0.095	0.458**	0.480**
	(0.289)	(0.098)	(0.206)	(0.202)
first_post_negemo	0.322**	-0.065	0.447***	0.411***
	(0.156)	(0.079)	(0.140)	(0.135)
total_6hours_difference	0.0109**	0.024***	0.003	0.004
	(0.004)	(0.004)	(0.004)	(0.004)
1.local_ol_author		-1.119		-0.426
		(1.652)		(1.103)
1.global_ol_author			3.000**	3.698***
			(1.469)	(1.285)
Constant	-5.77***	-1.033**	-4.109***	-3.660***
	(1.581)	(0.441)	(1.125)	(1.062)
Month	Included	Included	Included	Included
Time_of_day	Included	Included	Included	Included
Weekend	Included	Included	Included	Included
Observations	3,037	3,037	3,037	3,037
LR Chi2	162.77***	182.39***	174.61***	190.41***
Log-likelihood	-2246.573	-2222.722	-2232.436	-2214.544

3.5.4 Theoretical Contributions

This study examines the roles of opinion leaders in online health discussions on COVID-related mental health issues. Our study helps us better understand online health discussions by studying their content and contributors. In addition, it differentiates opinion leaders based on their connection in the online social network, content, and sentiment. Our study finds that global opinion leaders are more effective than local opinion leaders in the online community in terms of encouraging others to participate and get engaged with the community content. In addition, our sentiment analysis results help in understanding the importance of emotions, particularly negative ones, in elevating online mental health-related discussion, which could help in revealing individuals' disease-related fears, negative thoughts and beliefs about health problems, and attitudes toward medication and treatment.

Moreover, our study helps us better understand Reddit social website and explore the factors that affect its participation. This study finds that some opinion leaders foster Reddit thread participation. In addition, it finds that recently posted threads in a particular subreddit tend to have more active participation compared with older threads. Thus, this study helps to fill the gap in the literature about this rapidly growing yet less studied and explored platform compared with other popular platforms like Facebook and Twitter.

3.5.5 Practical Implications

This study sheds the light on the important role opinion leaders, particularly global opinion leaders, play in online health communities. Thus, social networking website managers could use mechanisms to identify influential individuals and employ the right people to encourage members' activity or get their attention to idle threads that require additional engagement.

Public health campaigns can be held online by opinion leaders in online health communities to spread health-related awareness regarding important issues and help combat health-related fake news dissemination due to their influence on others (Melchior & Oliveira, 2022). Furthermore, opinion leaders can assist healthcare providers in spreading health literacy and encouraging healthy lifestyles by sharing their health habits with others on social media and being role models that affect others' behavior and attitudes (L. Chen et al., 2019). For instance, opinion leaders could spread credible information about how to mitigate the spread of Covid virus and the importance of vaccination. For mental health problems, engagement in social websites is found to mitigate these types of problems by getting emotional support from others (Alonzo & Popescu, 2021). Since opinion leaders are found to increase others' participation, this increment will help increase social and emotional support, relieve stress or depression, and improve mental health.

3.6 Limitations and Future Research Directions

This study has some limitations that could be addressed in future research efforts. First, the study used historical data that were collected in a single time about Covid-related mental health discussions. Future research can leverage the power of longitudinal data to get a deeper insight into participation in the subreddit or community. Second, our study focused on the Reddit social platform. While more research is needed to explore Reddit, future research could use data from more than one social platform and compare the participation in each of them. Third, our study analyzed data from the mental health subreddit only. Future research could analyze the content of different subreddits and examine whether different subreddits have the same or different behaviors.

3.7 Conclusion

Our study extends the current understanding of online health discussions on the roles opinion leaders play to elevate participation in the discussion. Opinion leaders, particularly global opinion leaders, exert a significant influence on others and encouraged online community members' participation. In addition, emotions expressed by opinion leaders have an influence on others' behavior and participation. we employed social network analysis to identify local and global opinion leaders from the social network graphs of threads of posts. Moreover, we examined opinion leaders’ characteristics and content and their influence on subsequent participation in the subreddit thread. Our study is novel and contributes interesting findings that help to advance theory and leverage opinion leaders’ power and social media to spread health awareness and mitigate mental health problems.

REFERENCES

Aghili, G., & Lapointe, L. (2019). The Role of Online Communities in Vaccine Controversies.

Alazazi, M., & Ayaburi, E. W. (2019). Effect of Stigmatization and Privacy Concerns on Engagement in Virtual Health Communities.

Alazazi, M., & Wang, B. (2021). *Comparison of Voluntary versus Mandatory Vaccine Discussions in Online Health Communities: A Text Analytics Approach.* Paper presented at the Proceedings of the 54th Hawaii International Conference on System Sciences.

Alazazi, M., Wang, B., & Allan, T. (2020). Success factors of donation-based crowdfunding campaigns: A machine learning approach.

Alonzo, D., & Popescu, M. (2021). Utilizing social media platforms to promote mental health awareness and help seeking in underserved communities during the COVID-19 pandemic. *Journal of Education and Health Promotion, 10*.

Amith, M., Cohen, T., Cunningham, R., Savas, L. S., Smith, N., Cuccaro, P., . . . Tao, C. (2020). Mining HPV vaccine knowledge structures of young adults from Reddit using distributional semantics and pathfinder networks. *Cancer Control, 27*(1), 1073274819891442.

Andrews, V., Tonkin, E., Lancastle, D., & Kirk, M. (2014). Identifying the characteristics of nurse opinion leaders to aid the integration of genetics in nursing practice. *Journal of advanced nursing, 70*(11), 2598-2611.

Auxier, V., & Anderson, M. (2021). Social Media Use in 2021: A Majority of Americans Say They Use YouTube and Facebook, while Use of Instagram, Snapchat and TikTok Is Especially Common among Adults under 30. *Pew Research Center*.

Ayers, J. W., Westmaas, J. L., Leas, E. C., Benton, A., Chen, Y., Dredze, M., & Althouse, B. M. (2016). Leveraging big data to improve health awareness campaigns: a novel evaluation of the great American smokeout. *JMIR public health and surveillance, 2*(1), e16.

Bamakan, S. M. H., Nurgaliev, I., & Qu, Q. (2019). Opinion leader detection: A methodological review. *Expert Systems with Applications, 115*, 200-222.

Baumgartner, J., Zannettou, S., Keegan, B., Squire, M., & Blackburn, J. (2020). *The pushshift reddit dataset.* Paper presented at the Proceedings of the International AAAI Conference on Web and Social Media.

Bello-Orgaz, G., Hernandez-Castro, J., & Camacho, D. (2017). Detecting discussion communities on vaccination in twitter. *Future Generation Computer Systems, 66*, 125-136.

Bergström, A., & Jervelycke Belfrage, M. (2018). News in social media: Incidental consumption and the role of opinion leaders. *Digital Journalism, 6*(5), 583-598.

Biswas, B., Mukhopadhyay, A., & Gupta, G. (2018). *" Leadership in Action: How Top Hackers Behave" A Big-Data Approach with Text-Mining and Sentiment Analysis.* Paper presented at the Proceedings of the 51st Hawaii International Conference on System Sciences.

Biswas, B., Sengupta, P., & Chatterjee, D. (2020). Examining the determinants of the count of customer reviews in peer-to-peer home-sharing platforms using clustering and count regression techniques. *Decision Support Systems*, 113324.

Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., & Van Bavel, J. J. (2017). Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences, 114*(28), 7313-7318.

Brett, E. I., Stevens, E. M., Wagener, T. L., Leavens, E. L., Morgan, T. L., Cotton, W. D., & Hébert, E. T. (2019). A content analysis of JUUL discussions on social media: Using Reddit to understand patterns and perceptions of JUUL use. *Drug and alcohol dependence, 194*, 358-362.

Brunson, E. K. (2013). The impact of social networks on parents' vaccination decisions. *Pediatrics, 131*(5), e1397-e1404.

Buntain, C., & Golbeck, J. (2014). *Identifying social roles in reddit using network structure.* Paper presented at the Proceedings of the 23rd international conference on world wide web.

Buntinx-Krieg, T., Caravaglio, J., Domozych, R., & Dellavalle, R. P. (2017). Dermatology on Reddit: elucidating trends in dermatologic communications on the world wide web. *Dermatology online journal, 23*(7).

Burke-Garcia, A., Berry, C. N., Kreps, G. L., & Wright, K. B. (2017). *The power & perspective of mommy bloggers: Formative research with social media opinion leaders about HPV vaccination.* Paper presented at the Proceedings of the 50th Hawaii International Conference on System Sciences.

Burt, R. S. (1999). The social capital of opinion leaders. *The Annals of the American Academy of Political and Social Science, 566*(1), 37-54.

Carron-Arthur, B., Ali, K., Cunningham, J. A., & Griffiths, K. M. (2015). From help-seekers to influential users: a systematic review of participation styles in online health communities. *Journal of Medical Internet Research, 17*(12), e271.

Casaló, L. V., Flavián, C., & Ibáñez-Sánchez, S. (2018). Influencers on Instagram: Antecedents and consequences of opinion leadership. *Journal of Business Research*.

Cassis, C. (2019, January 23, 2019). Health & fitness on Reddit. Retrieved from https://www.redditinc.com/blog/health-fitness-on-reddit/

CDC. (2020). What are the benefits of flu vaccination? Retrieved from https://www.cdc.gov/flu/prevent/vaccine-benefits.htm

Chawla, S., & Mehrotra, M. (2021). Impact of emotions in social media content diffusion. *Informatica, 45*(6).

Chen, F. H. (2006). A susceptible-infected epidemic model with voluntary vaccinations. *Journal of mathematical biology, 53*(2), 253-272.

Chen, L., Baird, A., & Straub, D. (2019). Fostering participant health knowledge and attitudes: An econometric study of a chronic disease-focused online health community. *Journal of Management Information Systems, 36*(1), 194-229.

Chu, K.-H., Majmundar, A., Allem, J.-P., Soto, D. W., Cruz, T. B., & Unger, J. B. (2019). Tobacco use behaviors, attitudes, and demographic characteristics of tobacco opinion leaders and their followers: Twitter analysis. *Journal of Medical Internet Research, 21*(6), e12676.

Chun, H., Leem, B.-H., & Suh, H. (2021). Using text analytics to measure an effect of topics and sentiments on social-media engagement: focusing on Facebook fan page of Toyota. *International Journal of Engineering Business Management, 13*, 18479790211016268.

Cvitanic, T., Lee, B., Song, H. I., Fu, K., & Rosen, D. (2016). *LDA v. LSA: A comparison of two computational text analysis tools for the functional categorization of patents.* Paper presented at the International Conference on Case-Based Reasoning.

Daley, M. F., Narwaney, K. J., Shoup, J. A., Wagner, N. M., & Glanz, J. M. (2018). Addressing parents' vaccine concerns: a randomized trial of a social media intervention. *American journal of preventive medicine, 55*(1), 44-54.

Delnevo, G., Mirri, S., Monti, L., Prandi, C., Putra, M., Roccetti, M., . . . Sokol, R. J. (2018). *Patients reactions to non-invasive and invasive prenatal tests: a machine-based analysis from reddit posts.* Paper presented at the 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM).

Derksen, C., Serlachius, A., Petrie, K. J., & Dalbeth, N. (2017). "What say ye gout experts?" A content analysis of questions about gout posted on the social news website Reddit. *BMC musculoskeletal disorders, 18*(1), 488.

Donelson, C., Sutter, C., Pham, G. V., Narang, K., Wang, C., & Yun, J. T. (2021). Using a Machine Learning Methodology to Analyze Reddit Posts regarding Child Feeding Information. *Journal of Child and Family Studies*, 1-9.

Drum, D. A. (2009). Are toxic biometals destroying your children's future? *Biometals, 22*(5), 697-700.

Eghtesadi, M., & Florea, A. (2020). Facebook, Instagram, Reddit and TikTok: a proposal for health authorities to integrate popular social media platforms in contingency planning amid a global pandemic outbreak. *Canadian Journal of Public Health, 111*, 389-391.

El Naqa, I., & Murphy, M. J. (2015). *What is machine learning?* : Springer.

Elhadad, N., Zhang, S., Driscoll, P., & Brody, S. (2014). *Characterizing the sublanguage of online breast cancer forums for medications, symptoms, and emotions.* Paper presented at the AMIA Annual Symposium Proceedings.

Feldhege, J., Moessner, M., & Bauer, S. (2020). Who says what? Content and participation characteristics in an online depression community. *Journal of Affective Disorders, 263*, 521-527.

Ferrara, E., & Yang, Z. (2015). Quantifying the effect of sentiment on information diffusion in social media. *PeerJ Computer Science, 1*, e26.

Finney, L. J., & Iannotti, R. J. (2002). Message framing and mammography screening: A theory-driven intervention. *Behavioral Medicine, 28*(1), 5-14.

Flynn, L. R., Goldsmith, R. E., & Eastman, J. K. (1996). Opinion leaders and opinion seekers: Two new measurement scales. *Journal of the academy of marketing science, 24*(2), 137-147.

Fraga, B. S., da Silva, A. P. C., & Murai, F. (2018). *Online social networks in health care: a study of mental disorders on Reddit.* Paper presented at the 2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI).

Franke, F., Felfe, J., & Pundt, A. (2014). The impact of health-oriented leadership on follower health: Development and test of a new instrument measuring health-promoting leadership. *German Journal of Human Resource Management, 28*(1-2), 139-161.

Fukuda, E., Kokubo, S., Tanimoto, J., Wang, Z., Hagishima, A., & Ikegaya, N. (2014). Risk assessment for infectious disease and its impact on voluntary vaccination behavior in social networks. *Chaos, Solitons & Fractals, 68*, 1-9.

Galanakis, E., Jansen, A., Lopalco, P. L., & Giesecke, J. (2013). Ethics of mandatory vaccination for healthcare workers. *Eurosurveillance, 18*(45), 20627.

Gallagher, K. M., & Updegraff, J. A. (2012). Health message framing effects on attitudes, intentions, and behavior: a meta-analytic review. *Annals of behavioral medicine, 43*(1), 101-116.

Gefen, D., Endicott, J. E., Fresneda, J. E., Miller, J., & Larsen, K. R. (2017). A guide to text analysis with latent semantic analysis in R with annotated code: Studying online reviews and the stack exchange community. *Communications of the Association for Information Systems, 41*(1), 21.

Gesser-Edelsburg, A., Walter, N., Shir-Raz, Y., & Green, M. S. (2015). Voluntary or mandatory? The valence framing effect of attitudes regarding HPV vaccination. *Journal of health communication, 20*(11), 1287-1293.

Glanz, J. M., Wagner, N. M., Narwaney, K. J., Kraus, C. R., Shoup, J. A., Xu, S., . . . Daley, M. F. (2017). Web-based social media intervention to increase vaccine acceptance: a randomized controlled trial. *Pediatrics, 140*(6), e20171117.

Guimaraes, A., Balalau, O., Terolli, E., & Weikum, G. (2019). *Analyzing the traits and anomalies of political discussions on reddit.* Paper presented at the Proceedings of the International AAAI Conference on Web and Social Media.

Hajli, M. N., Sims, J., Featherman, M., & Love, P. E. (2015). Credibility of information in online communities. *Journal of Strategic Marketing, 23*(3), 238-253.

Haralabopoulos, G., Anagnostopoulos, I., & Zeadally, S. (2015). Lifespan and propagation of information in On-line Social Networks: A case study based on Reddit. *Journal of network and computer applications, 56*, 88-100.

Haverkate, M., D'Ancona, F., Giambi, C., Johansen, K., Lopalco, P. L., Cozza, V., & Appelgren, E. (2012). Mandatory and recommended vaccination in the EU, Iceland and Norway: results of the VENICE 2010 survey on the ways of implementing national vaccination programmes. *Eurosurveillance, 17*(22), 20183.

Hawi, N. S., & Samaha, M. (2017). The relations among social media addiction, self-esteem, and life satisfaction in university students. *Social Science Computer Review, 35*(5), 576-586.

Hemmings-Jarrett, K., Jarrett, J., & Blake, M. B. (2018). *(WKSP) Sentiment Analysis of Twitter Samples That Differentiates Impact of User Participation Levels.* Paper presented at the 2018 IEEE International Conference on Cognitive Computing (ICCC).

Hossain, M. A. (2019). Effects of uses and gratifications on social media use. *PSU Research Review*.

Hou, L. (2022). Network versus content: The effectiveness in identifying opinion leaders in an online social network with empirical evaluation. *Physica A: Statistical Mechanics and its Applications, 592*, 126879.

Huang, X., Smith, M. C., Paul, M. J., Ryzhkov, D., Quinn, S. C., Broniatowski, D. A., & Dredze, M. (2017). *Examining patterns of influenza vaccination in social media.* Paper presented at the Workshops at the Thirty-First AAAI Conference on Artificial Intelligence.

Huh, J., McDonald, D. W., Hartzler, A., & Pratt, W. (2013). *Patient moderator interaction in online health communities.* Paper presented at the AMIA Annual Symposium Proceedings.

Igawa, R., Almeida, A., & Zarpelão, B. (2015). *Recognition of compromised accounts on twitter.* Paper presented at the Anais Principais do XI Simpósio Brasileiro de Sistemas de Informação.

James W. Pennebaker, R. L. B., & Kayla Jordan, a. K. B. (2015). The development and psychometric properties of LIWC2015. *Austin, TX: University of Texas at Austin*.

Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., & Zhao, L. (2019). Latent Dirichlet Allocation (LDA) and Topic modeling: models, applications, a survey. *Multimedia Tools and Applications, 78*(11), 15169-15211.

Jones, Q., Ravid, G., & Rafaeli, S. (2004). Information overload and the message dynamics of online interaction spaces: A theoretical model and empirical exploration. *Information systems research, 15*(2), 194-210.

Joyce, E., & Kraut, R. E. (2006). Predicting continued participation in newsgroups. *Journal of Computer-Mediated Communication, 11*(3), 723-747.

Kang, G. J., Ewing-Nelson, S. R., Mackey, L., Schlitt, J. T., Marathe, A., Abbas, K. M., & Swarup, S. (2017). Semantic network analysis of vaccine sentiment in online social media. *Vaccine, 35*(29), 3621-3638.

Karlsen, R. (2015). Followers are opinion leaders: The role of people in the flow of political communication on and beyond social networking sites. *European journal of communication, 30*(3), 301-318.

Katz, E. (1957). The two-step flow of communication: An up-to-date report on an hypothesis. *Public opinion quarterly, 21*(1), 61-78.

Katz, E., & Lazarsfeld, P. (1955). Personal Influence New York. *NY: The Free Press.[Google Scholar]*.

Keib, K., Espina, C., Lee, Y.-I., Wojdynski, B. W., Choi, D., & Bang, H. (2018). Picture this: The influence of emotionally valenced images, on attention, selection, and sharing of social media news. *Media Psychology, 21*(2), 202-221.

Kim, E., Hou, J., Han, J. Y., & Himelboim, I. (2016). Predicting retweeting behavior on breast cancer social networks: Network and content characteristics. *Journal of health communication, 21*(4), 479-486.

Kim, E., Scheufele, D. A., Han, J. Y., & Shah, D. (2017). Opinion leaders in online cancer support groups: An investigation of their antecedents and consequences. *Health Communication, 32*(2), 142-151.

Kim, H. S., Lee, S., Cappella, J. N., Vera, L., & Emery, S. (2013). Content characteristics driving the diffusion of antismoking messages: Implications for cancer prevention in the emerging public communication environment. *Journal of the National Cancer Institute Monographs, 2013*(47), 182-187.

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*, 153-165.

Knoke, D., & Yang, S. (2019). *Social network analysis*: Sage Publications.

Kühl, N., Mühlthaler, M., & Goutier, M. (2018). *Automatically quantifying customer need tweets: Towards a supervised machine learning approach.* Paper presented at the Proceedings of the 51st Hawaii International Conference on System Sciences.

Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling* (Vol. 26): Springer.
Lama, Y., Hu, D., Jamison, A., Quinn, S. C., & Broniatowski, D. A. (2019). Characterizing trends in human papillomavirus vaccine discourse on Reddit (2007-2015): an observational study. *JMIR public health and surveillance, 5*(1), e12480.
Lamirán-Palomares, J. M., Baviera, T., & Baviera-Puig, A. (2019). Identifying Opinion Leaders on Twitter during sporting events: lessons from a case study. *Social Sciences, 8*(5), 141.
Langley, D. J., Wijn, R., Epskamp, S., & Van Bork, R. (2015). Should I get that jab? Exploring influence to encourage vaccination via online social media.
Lazarsfeld, P. F., Berelson, B., & Gaudet, H. (1944). The people's choice.
Little, K., Goodridge, S., Lewis, H., Lingard, S., Din, S., Tidley, M., . . . Hayes, S. (2015). Occupational vaccination of health care workers: uptake, attitudes and potential solutions. *Public Health, 129*(6), 755-762.
Liu, Q. B., Liu, X., & Guo, X. (2020). The Effects of Participating in a Physician-Driven Online Health Community in Managing Chronic Disease: Evidence from Two Natural Experiments. *MIS Quarterly, 44*(1).
Liu, Y., Gu, Z., Ko, T. H., & Liu, J. (2018). Identifying key opinion leaders in social media via modality-consistent harmonized discriminant embedding. *IEEE transactions on cybernetics, 50*(2), 717-728.
Liu, Y., Wang, Y., & Zhang, J. (2012). *New machine learning algorithm: Random forest.* Paper presented at the Information Computing and Applications: Third International Conference, ICICA 2012, Chengde, China, September 14-16, 2012. Proceedings 3.
Liu, Y., & Yin, Z. (2020). Understanding Weight Loss via Online Discussions: Content Analysis of Reddit Posts Using Topic Modeling and Word Clustering Techniques. *Journal of Medical Internet Research, 22*(6), e13745.
Love, B., Himelboim, I., Holton, A., & Stewart, K. (2013). Twitter as a source of vaccination information: content drivers and what they are saying. *American journal of infection control, 41*(6), 568-570.
Lu, Y., Wu, Y., Liu, J., Li, J., & Zhang, P. (2017). Understanding health care social media use from different stakeholder perspectives: a content analysis of an online health community. *Journal of Medical Internet Research, 19*(4), e109.
Luo, L., Yang, Y., Chen, Z., & Wei, Y. (2018). Identifying opinion leaders with improved weighted LeaderRank in online learning communities. *International Journal of Performability Engineering, 14*(2), 193.
Mahesh, B. (2020). Machine learning algorithms-a review. *International Journal of Science and Research (IJSR).[Internet], 9*, 381-386.
Margolis, M. A., Brewer, N. T., Shah, P. D., Calo, W. A., & Gilkey, M. B. (2019). Stories about HPV vaccine in social media, traditional media, and conversations. *Preventive medicine, 118*, 251-256.
Massey, P. M., Kearney, M. D., Hauer, M. K., Selvan, P., Koku, E., & Leader, A. E. (2020). Dimensions of misinformation about the HPV vaccine on Instagram: Content and network analysis of social media characteristics. *Journal of Medical Internet Research, 22*(12), e21451.
Mayshak, R., Sharman, S. J., Zinkiewicz, L., & Hayley, A. (2017). The influence of empathy and self-presentation on engagement with social networking website posts. *Computers in Human Behavior, 71*, 362-377.

Melchior, C., & Oliveira, M. (2022). Health-related fake news on social media platforms: A systematic literature review. *new media & society, 24*(6), 1500-1522.
Melton, C. A., Olusanya, O. A., Ammar, N., & Shaban-Nejad, A. (2021). Public sentiment analysis and topic modeling regarding COVID-19 vaccines on the Reddit social media platform: A call to action for strengthening vaccine confidence. *Journal of Infection and Public Health, 14*(10), 1505-1512.
Mhamdi, C., Al-Emran, M., & Salloum, S. A. (2018). Text mining and analytics: A case study from news channels posts on Facebook. In *Intelligent Natural Language Processing: Trends and Applications* (pp. 399-415): Springer.
Minami, M., Lennert-Cody, C. E., Gao, W., & Román-Verdesoto, M. (2007). Modeling shark bycatch: the zero-inflated negative binomial regression model with smoothing. *Fisheries Research, 84*(2), 210-221.
Moghimbeigi, A., Eshraghian, M. R., Mohammad, K., & Mcardle, B. (2008). Multilevel zero-inflated negative binomial regression modeling for over-dispersed count data with extra zeros. *Journal of Applied Statistics, 35*(10), 1193-1202.
Mohamad, E., Ahmad, A. L., Salleh, S. M., & Sulaiman, U. K. W. (2017). *Breaking the traditional communication flow: Exploration of social media opinion leaders in health.* Paper presented at the SHS Web of Conferences.
Mollema, L., Harmsen, I. A., Broekhuizen, E., Clijnk, R., De Melker, H., Paulussen, T., . . . Das, E. (2015). Disease detection or public opinion reflection? Content analysis of tweets, other social media, and online newspapers during the measles outbreak in The Netherlands in 2013. *Journal of Medical Internet Research, 17*(5), e128.
Momtaz, N. J., Aghaie, A., & Alizadeh, S. (2011). Identifying opinion leaders for marketing by analyzing online social networks. *International Journal of Virtual Communities and Social Networking (IJVCSN), 3*(3), 19-34.
Monselise, M., Chang, C.-H., Ferreira, G., Yang, R., & Yang, C. C. (2021). Topics and sentiments of public concerns regarding COVID-19 vaccines: social media trend analysis. *Journal of Medical Internet Research, 23*(10), e30765.
Mukherjee, S., Weikum, G., & Danescu-Niculescu-Mizil, C. (2014). *People on drugs: credibility of user statements in health communities.* Paper presented at the Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining.
Myneni, S., Fujimoto, K., Cobb, N., & Cohen, T. (2015). Content-driven analysis of an online community for smoking cessation: integration of qualitative techniques, automated text analysis, and affiliation networks. *American journal of public health, 105*(6), 1206-1212.
Nisbet, E. C. (2006). The engagement model of opinion leadership: Testing validity within a European context. *International Journal of Public Opinion Research, 18*(1), 3-30.
Nunes, R. H., Ferreira, J. B., de Freitas, A. S., & Ramos, F. L. (2018). The effects of social media opinion leaders' recommendations on followers' intention to buy. *Revista Brasileira de Gestão de Negócios, 20*(1), 57-73.
Nyhan, B., & Reifler, J. (2015). Does correcting myths about the flu vaccine work? An experimental evaluation of the effects of corrective information. *Vaccine, 33*(3), 459-464.
Oueslati, W., Arrami, S., Dhouioui, Z., & Massaabi, M. (2021). Opinion leaders' detection in dynamic social networks. *Concurrency and Computation: Practice and Experience, 33*(1), e5692.
Park, A., Conway, M., & Chen, A. T. (2018). Examining thematic similarity, difference, and

membership in three online mental health communities from Reddit: a text mining and visualization approach. *Computers in Human Behavior, 78*, 98-112.
Pennebaker, J. W., Booth, R. J., Boyd, R. L., & Francis, M. E. (2015). Linguistic inquiry and word count: Liwc 2015 [computer software]. pennebaker conglomerates. In: Inc.
Pirina, I., & Çöltekin, Ç. (2018). *Identifying depression on reddit: The effect of training data.* Paper presented at the Proceedings of the 2018 EMNLP Workshop SMM4H: The 3rd Social Media Mining for Health Applications Workshop & Shared Task.
Praveen, S., Ittamalla, R., & Deepak, G. (2021). Analyzing the attitude of Indian citizens towards COVID-19 vaccine–A text analytics study. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews, 15*(2), 595-599.
Pulido, C. M., Ruiz-Eugenio, L., Redondo-Sama, G., & Villarejo-Carballido, B. (2020). A new application of social impact in social media for overcoming fake news in health. *International journal of environmental research and public health, 17*(7), 2430.
Puri, N., Coomes, E. A., Haghbayan, H., & Gunaratne, K. (2020). Social media and vaccine hesitancy: new updates for the era of COVID-19 and globalized infectious diseases. *Human Vaccines & Immunotherapeutics*, 1-8.
Rehman, A. U., Jiang, A., Rehman, A., Paul, A., & Sadiq, M. T. (2020). Identification and role of opinion leaders in information diffusion for online discussion network. *Journal of Ambient Intelligence and Humanized Computing*, 1-13.
Rhidenour, K. B., Blackburn, K., Barrett, A. K., & Taylor, S. (2021). Mediating Medical Marijuana: Exploring How Veterans Discuss Their Stigmatized Substance Use on Reddit. *Health Communication*, 1-11.
Riquelme, F., Gonzalez-Cantergiani, P., Hans, D., Villarroel, R., & Munoz, R. (2019). Identifying opinion leaders on social networks through milestones definition. *IEEE Access, 7*, 75670-75677.
Rogers, E. M., & Cartano, D. G. (1962). Methods of measuring opinion leadership. *Public opinion quarterly*, 435-441.
Rothman, A. J., & Salovey, P. (1997). Shaping perceptions to motivate healthy behavior: the role of message framing. *Psychological bulletin, 121*(1), 3.
Rout, J. K., Choo, K.-K. R., Dash, A. K., Bakshi, S., Jena, S. K., & Williams, K. L. (2018). A model for sentiment and emotion analysis of unstructured social media text. *Electronic Commerce Research, 18*(1), 181-199.
Salloum, S. A., Al-Emran, M., Monem, A. A., & Shaalan, K. (2017). A survey of text mining in social media: facebook and twitter perspectives. *Adv. Sci. Technol. Eng. Syst. J, 2*(1), 127-133.
Salloum, S. A., Al-Emran, M., & Shaalan, K. (2017). Mining social media text: extracting knowledge from Facebook. *International Journal of Computing and Digital Systems, 6*(02), 73-81.
Schäfer, M. S., & Taddicken, M. (2015). Opinion Leadership| Mediatized Opinion Leaders: New Patterns of Opinion Leadership in New Media Environments? *International Journal of Communication, 9*, 22.
Serrat, O. (2017). Social network analysis. In *Knowledge solutions* (pp. 39-43): Springer.
Shepherd, A., Sanders, C., Doyle, M., & Shaw, J. (2015). Using social media for support and feedback by mental health service users: thematic analysis of a twitter conversation. *BMC psychiatry, 15*(1), 29.

Shi, J., & Salmon, C. T. (2018). Identifying opinion leaders to promote organ donation on social media: network study. *Journal of Medical Internet Research, 20*(1), e7.

Shoup, J. A., Wagner, N. M., Kraus, C. R., Narwaney, K. J., Goddard, K. S., & Glanz, J. M. (2015). Development of an interactive social media tool for parents with concerns about vaccines. *Health Education & Behavior, 42*(3), 302-312.

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), 421-455.

Singer, P., Flöck, F., Meinhart, C., Zeitfogel, E., & Strohmaier, M. (2014). *Evolution of reddit: from the front page of the internet to a self-referential community?* Paper presented at the Proceedings of the 23rd international conference on world wide web.

Smith, S. M., & Petty, R. E. (1996). Message framing and persuasion: A message processing analysis. *Personality and Social Psychology Bulletin, 22*(3), 257-268.

Song, S. Y., Cho, E., & Kim, Y.-K. (2017). Personality factors and flow affecting opinion leadership in social media. *Personality and Individual Differences, 114*, 16-23.

Song, X., Chi, Y., Hino, K., & Tseng, B. (2007). *Identifying opinion leaders in the blogosphere.* Paper presented at the Proceedings of the sixteenth ACM conference on Conference on information and knowledge management.

Sowles, S. J., McLeary, M., Optican, A., Cahn, E., Krauss, M. J., Fitzsimmons-Craft, E. E., . . . Cavazos-Rehg, P. A. (2018). A content analysis of an online pro-eating disorder community on Reddit. *Body image, 24*, 137-144.

Stieglitz, S., & Dang-Xuan, L. (2013). Emotions and information diffusion in social media—sentiment of microblogs and sharing behavior. *Journal of Management Information Systems, 29*(4), 217-248.

Sun, G., & Bin, S. (2018). A new opinion leaders detecting algorithm in multi-relationship online social networks. *Multimedia Tools and Applications, 77*(4), 4295-4307.

Taunk, K., De, S., Verma, S., & Swetapadma, A. (2019). *A brief review of nearest neighbor algorithm for learning and classification.* Paper presented at the 2019 International Conference on Intelligent Computing and Control Systems (ICCS).

Thelwall, M. (2018). Can social news websites pay for content and curation? The SteemIt cryptocurrency model. *Journal of Information Science, 44*(6), 736-751.

Trucchi, C., Paganino, C., Orsi, A., De Florentiis, D., & Ansaldi, F. (2015). Influenza vaccination in the elderly: why are the overall benefits still hotly debated? *Journal of preventive medicine and hygiene, 56*(1), E37.

Van't Riet, J., Cox, A. D., Cox, D., Zimet, G. D., De Bruijn, G.-J., Van den Putte, B., . . . Ruiter, R. A. (2016). Does perceived risk influence the effects of message framing? Revisiting the link between prospect theory and message framing. *Health psychology review, 10*(4), 447-459.

Vijayarani, S., Ilamathi, M. J., & Nithya, M. (2015). Preprocessing techniques for text mining-an overview. *International Journal of Computer Science & Communication Networks, 5*(1), 7-16.

Wang, T., Brede, M., Ianni, A., & Mentzakis, E. (2018). Social interactions in online eating disorder communities: A network perspective. *PloS one, 13*(7), e0200800.

Wang, X., Chen, L., Shi, J., & Peng, T.-Q. (2019). What makes cancer information viral on social media? *Computers in Human Behavior, 93*, 149-156.

Wang, X., & Lee, E. W. (2020). Negative emotions shape the diffusion of cancer tweets: toward an integrated social network–text analytics approach. *Internet Research*.

Wang, X., Zhao, K., & Street, N. (2017). Analyzing and predicting user participations in online health communities: a social support perspective. *Journal of Medical Internet Research, 19*(4), e6834.

Wang, Y., & Fikis, D. J. (2019). Common core state standards on Twitter: Public sentiment and opinion leaders. *Educational Policy, 33*(4), 650-683.

Watson, J. C., Hadler, S. C., Dykewicz, C. A., Reef, S., & Phillips, L. (1998). *Measles, Mumps, and Rubella-Vaccine Use and Strategies for Elimination of Measles, Rubella, and Congenital Rubella Syndrome and Control of Mumps: Recommendations of the Advisory Committee on Immunization Practices (ACIP). Vol. 47/No. RR-8*. Retrieved from

Williams, T., & Betak, J. (2018). A Comparison of LSA and LDA for the Analysis of Railroad Accident Text. *Procedia computer science, 130*, 98-102.

Willis, E. (2018). Applying the health belief model to medication adherence: the role of online health communities and peer reviews. *Journal of health communication, 23*(8), 743-750.

Wimmer, H., Yoon, V. Y., & Sugumaran, V. (2016). A multi-agent system to support evidence based medicine and clinical decision making via data sharing and data privacy. *Decision Support Systems, 88*, 51-66.

Winkelmann, R. (2000). Seemingly unrelated negative binomial regression. *Oxford Bulletin of Economics and Statistics, 62*(4), 553-560.

Winter, S., & Neubaum, G. (2016). Examining characteristics of opinion leaders in social media: A motivational approach. *Social Media+ Society, 2*(3), 2056305116665858.

Xiong, Y., Cheng, Z., Liang, E., & Wu, Y. (2018). Accumulation mechanism of opinion leaders' social interaction ties in virtual communities: Empirical evidence from China. *Computers in Human Behavior, 82*, 81-93.

Xu, Q., Yu, N., & Song, Y. (2018). User engagement in public discourse on genetically modified organisms: The role of opinion leaders on social media. *Science communication, 40*(6), 691-717.

Xu, W. W., Chiu, I.-H., Chen, Y., & Mukherjee, T. (2015). Twitter hashtags for health: applying network and content analyses to understand the health knowledge sharing in a Twitter-based community of practice. *Quality & Quantity, 49*(4), 1361-1380.

Xu, X., Li, Z., Wang, R., & Zhao, L. (2021). Analysis of the evolution of user emotion and opinion leaders' information dissemination behavior in the knowledge Q&A community during COVID-19. *International journal of environmental research and public health, 18*(22), 12252.

Yan, Z., Wang, T., Chen, Y., & Zhang, H. (2016). Knowledge sharing in online health communities: A social exchange theory perspective. *Information & management, 53*(5), 643-653.

Yang, L., Qiao, Y., Liu, Z., Ma, J., & Li, X. (2018). Identifying opinion leader nodes in online social networks with a new closeness evaluation algorithm. *Soft Computing, 22*(2), 453-464.

Yin, F., Xia, X., Song, N., Zhu, L., & Wu, J. (2020). Quantify the role of superspreaders-opinion leaders-on COVID-19 information propagation in the Chinese Sina-microblog. *PloS one, 15*(6), e0234023.

Yoo, K.-H., Gretzel, U., & Zach, F. (2011). Travel opinion leaders and seekers. In *Information and Communication Technologies in Tourism 2011* (pp. 525-535): Springer.

Yue, L., Chen, W., Li, X., Zuo, W., & Yin, M. (2019). A survey of sentiment analysis in social media. *Knowledge and Information Systems, 60*(2), 617-663.

Yun, G. W., Morin, D., Park, S., Joa, C. Y., Labbe, B., Lim, J., . . . Hyun, D. (2016). Social media and flu: Media Twitter accounts as agenda setters. *International journal of medical informatics, 91*, 67-73.

Zhai, Z., Xu, H., & Jia, P. (2008). *Identifying opinion leaders in BBS.* Paper presented at the 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology.

Zhan, Y., Zhang, Z., Okamoto, J. M., Zeng, D. D., & Leischow, S. J. (2019). Underage juul use patterns: Content analysis of reddit messages. *Journal of Medical Internet Research, 21*(9), e13038.

Zhang, X., Liu, S., Deng, Z., & Chen, X. (2017). Knowledge sharing motivations in online health communities: A comparative study of health professionals and normal users. *Computers in Human Behavior, 75*, 797-810.

Zhang, Y. (2012). *Support vector machine classification algorithm and its application.* Paper presented at the Information Computing and Applications: Third International Conference, ICICA 2012, Chengde, China, September 14-16, 2012. Proceedings, Part II 3.

Zhao, K., Yen, J., Greer, G., Qiu, B., Mitra, P., & Portier, K. (2014). Finding influential users of online health communities: a new metric based on sentiment influence. *Journal of the American Medical Informatics Association, 21*(e2), e212-e218.

Zhao, Y., Kou, G., Peng, Y., & Chen, Y. (2018). Understanding influence power of opinion leaders in e-commerce networks: An opinion dynamics theory perspective. *Information Sciences, 426*, 131-147.

Zheng, K., Li, A., & Farzan, R. (2018). *Exploration of online health support groups through the lens of sentiment analysis.* Paper presented at the International Conference on Information.

Zhou, S., Barnes, L., McCormick, H., & Cano, M. B. (2021). Social media influencers' narrative strategies to create eWOM: A theoretical contribution. *International Journal of Information Management, 59*, 102293.

Zunic, A., Corcoran, P., & Spasic, I. (2020). Sentiment analysis in health and well-being: systematic review. *JMIR medical informatics, 8*(1), e16023.

APPENDIX A

APPENDIX A

Table 17. Term Association Analysis Results for First Thread Posts of Flu Vaccine-Related Discussions.

Term 1	Term 2	Correlation
flu	shot	0.64
flu	get	0.59
flu	can	0.57
flu	prevent	0.57
flu	even	0.56
flu	just	0.55
flu	still	0.55
flu	one	0.52
flu	question	0.51
flu	peopl	0.50
flu	protect	0.46
flu	sinc	0.44
flu	happen	0.44
flu	right	0.44
flu	around	0.43
flu	take	0.42
flu	like	0.41
flu	time	0.40
flu	think	0.38
flu	never	0.36
flu	year	0.35
flu	cough	0.35
flu	realli	0.35
flu	week	0.35
flu	help	0.35
flu	feel	0.34
flu	lot	0.34
flu	said	0.34
flu	dont	0.34
flu	person	0.34
flu	infect	0.33
flu	last	0.32
flu	read	0.32
flu	sick	0.32

Table 17, cont.

flu	thank	0.31
flu	virus	0.31
flu	first	0.29
flu	make	0.29
flu	much	0.29
flu	know	0.29
flu	way	0.28
flu	will	0.28
flu	say	0.27
flu	babi	0.27
flu	possibl	0.27
flu	strain	0.27
flu	infant	0.27
flu	noth	0.26
flu	tri	0.26
flu	also	0.24
flu	got	0.24
flu	start	0.23
flu	want	0.22
flu	sever	0.22
flu	Can’t	0.22
flu	back	0.21
flu	believ	0.20

APPENDIX B

APPENDIX B

Table 18. Term Association Analysis Results for First Thread Posts of MMR Vaccine-Related Discussions.

Term 1	Term 2	Correlation
mmr	control	0.96
mmr	support	0.94
mmr	data	0.92
mmr	conclus	0.91
mmr	born	0.90
mmr	asd	0.89
mmr	autism	0.88
mmr	development	0.87
mmr	disord	0.87
mmr	increas	0.86
mmr	evid	0.86
mmr	incid	0.85
mmr	associ	0.84
mmr	children	0.83
mmr	compar	0.83
mmr	age	0.82
mmr	committe	0.82
mmr	diagnos	0.82
mmr	regress	0.81
mmr	group	0.80
mmr	among	0.80
mmr	causal	0.80
mmr	preval	0.80
mmr	measlesmumpsrubella	0.78
mmr	provid	0.78
mmr	casecontrol	0.77
mmr	signific	0.76
mmr	studi	0.75
mmr	older	0.75
mmr	either	0.74
mmr	estim	0.74
mmr	period	0.73
mmr	without	0.72
mmr	record	0.72
mmr	gastrointestin	0.72
mmr	time	0.71
mmr	result	0.70
mmr	cohort	0.70
mmr	vaccin	0.69

Table 18, cont.

mmr	found	0.69
mmr	whether	0.68
mmr	general	0.68
mmr	base	0.68
mmr	hour	0.67
mmr	unit	0.67
mmr	correl	0.67
mmr	child	0.66
mmr	year	0.65
mmr	later	0.65
mmr	risk	0.65
mmr	relat	0.65
mmr	receiv	0.64
mmr	relationship	0.64
mmr	popul	0.63
mmr	persist	0.62
mmr	develop	0.61
mmr	nurs	0.61
mmr	total	0.60
mmr	epidemiolog	0.59
mmr	practic	0.59
mmr	differ	0.58
mmr	sinc	0.58
mmr	strong	0.58
mmr	recogn	0.58
mmr	rise	0.58
mmr	virus	0.57
mmr	research	0.57
mmr	three	0.56
mmr	review	0.56
mmr	diseas	0.55
mmr	respons	0.55
mmr	rubella	0.54
mmr	suggest	0.54
mmr	birth	0.53
mmr	syndrom	0.53
mmr	current	0.53
mmr	blood	0.52
mmr	look	0.52
mmr	within	0.51
mmr	exist	0.51
mmr	ten	0.51
mmr	major	0.51
mmr	decreas	0.50

Table 18, cont.

mmr	notabl	0.50
mmr	measl	0.49
mmr	initi	0.49
mmr	communic	0.49
mmr	antibodi	0.48
mmr	symptom	0.48
mmr	use	0.48
mmr	analysi	0.48
mmr	rate	0.48
mmr	safeti	0.48
mmr	administ	0.47
mmr	across	0.47
mmr	negat	0.46
mmr	upon	0.46
mmr	dose	0.45
mmr	immun	0.45
mmr	august	0.45
mmr	specif	0.44
mmr	furthermor	0.44
mmr	contain	0.44
mmr	problem	0.44
mmr	trigger	0.44
mmr	recommend	0.43
mmr	also	0.42
mmr	like	0.42
mmr	follow	0.42
mmr	earli	0.42
mmr	first	0.41
mmr	event	0.41
mmr	accept	0.41
mmr	expect	0.40
mmr	histori	0.40
mmr	center	0.40
mmr	shown	0.40
mmr	posit	0.39
mmr	includ	0.39
mmr	case	0.38
mmr	health	0.38
mmr	program	0.38
mmr	childhood	0.38
mmr	exposur	0.37
mmr	subgroup	0.37
mmr	receipt	0.37
mmr	status	0.37

Table 18, cont.

mmr	reaction	0.36
mmr	month	0.36
mmr	children	0.36
mmr	perform	0.36
mmr	post	0.36
mmr	link	0.36
mmr	autist	0.36
mmr	anoth	0.35
mmr	concern	0.34
mmr	parent	0.34
mmr	contrast	0.34
mmr	decis	0.34
mmr	infecti	0.34
mmr	loss	0.34
mmr	independ	0.34
mmr	find	0.33
mmr	caus	0.33
mmr	inform	0.33
mmr	occur	0.32
mmr	million	0.32
mmr	advers	0.32
mmr	effect	0.31
mmr	part	0.31
mmr	better	0.31
mmr	clear	0.31
mmr	larg	0.31
mmr	examin	0.31
mmr	region	0.31
mmr	mump	0.30
mmr	well	0.30
mmr	lead	0.30
mmr	high	0.30
mmr	state	0.30
mmr	level	0.29
mmr	life	0.29
mmr	mani	0.29
mmr	show	0.29
mmr	one	0.28
mmr	due	0.28
mmr	add	0.28
mmr	scientist	0.28
mmr	possibl	0.28
mmr	countri	0.27
mmr	serious	0.27

Table 18, cont.

mmr	normal	0.27
mmr	medicin	0.27
mmr	april	0.26
mmr	fine	0.26
mmr	healthiest	0.26
mmr	number	0.26
mmr	offici	0.26
mmr	new	0.25
mmr	two	0.25
mmr	explain	0.25
mmr	public	0.25
mmr	addit	0.25
mmr	ban	0.25
mmr	govern	0.25
mmr	japan	0.25
mmr	remov	0.25
mmr	neurolog	0.25
mmr	combin	0.24
mmr	industri	0.24
mmr	second	0.24
mmr	factor	0.24
mmr	report	0.24
mmr	patient	0.24
mmr	potenti	0.24
mmr	test	0.23
mmr	peopl	0.23
mmr	never	0.23
mmr	suffer	0.23
mmr	global	0.23
mmr	japanes	0.23
mmr	mandat	0.23
mmr	enceph	0.23
mmr	disabl	0.23
mmr	expert	0.23
mmr	registri	0.23
mmr	person	0.22
mmr	separ	0.22
mmr	offer	0.22
mmr	institut	0.22
mmr	prevent	0.22
mmr	regard	0.22
mmr	author	0.22
mmr	coverag	0.22
mmr	requir	0.21

Table 18, cont.

mmr	side	0.21
mmr	articl	0.21
mmr	given	0.21
mmr	febril	0.21
mmr	nation	0.21
mmr	observ	0.21
mmr	view	0.21
mmr	worth	0.21
mmr	happen	0.20
mmr	compulsori	0.20
mmr	dean	0.20
mmr	repres	0.20

APPENDIX C

APPENDIX C

Table 19. Regression Results for the 5% Opinion Leaders

Variables	Model 1	Model 2	Model 3	Model 4
Negative Binomial Part				
first_post_posemo	0.0204	0.0392	0.0199	0.0186
	(0.0318)	(0.0318)	(0.0321)	(0.0326)
first_post_negemo	-0.00151	-0.00963	0.0132	0.000822
	(0.0238)	(0.0238)	(0.0244)	(0.0239)
total_6hours_difference	-0.0162***	-0.0146***	-0.0145***	-0.0133***
	(0.00176)	(0.00178)	(0.00176)	(0.00169)
1.local_ol_author	0.0650	-0.532**		-0.765***
	(0.157)	(0.263)		(0.261)
Lag1_last_local_ol_count		0.721***		0.337
		(0.274)		(0.298)
Lag1_avg_posemo_local_opl		-0.0144		0.0885
		(0.0596)		(0.0789)
Lag1_avg_negemo_local_opl		0.0154		-0.0300
		(0.0449)		(0.0726)
1.global_ol_author			-1.165**	0.752
			(0.458)	(0.698)
Lag1_last_global_ol_count			0.737***	0.681***
			(0.198)	(0.223)
Lag1_avg_posemo_global_opl			-0.0704	-0.129*
			(0.0504)	(0.0739)
Lag1_avg_negemo_global_opl			0.0557	0.0372
			(0.0393)	(0.0652)
Constant	0.0650	0.0205	0.0775	0.0265
	(0.157)	(0.157)	(0.163)	(0.159)
Month	Included	Included	Included	Included
Time_of_day	Included	Included	Included	Included
Weekend	Included	Included	Included	Included
Zero-Inflated Part				
first_post_posemo	1.043***	1.038***	0.796***	1.035***

Table 19, cont.

	(0.289)	(0.301)	(0.251)	(0.304)
first_post_negemo	0.322**	0.314*	0.420***	0.378**
	(0.156)	(0.161)	(0.148)	(0.188)
total_6hours_difference	0.0109**	0.0117***	0.00523	0.0126**
	(0.00446)	(0.00454)	(0.00371)	(0.00528)
1.local_ol_author		0.312		-8.189**
		(1.357)		(3.210)
1.global_ol_author			-16.31	8.650***
			(2,025)	(2.575)
Constant	-5.773***	-5.692***	-4.691***	-5.895***
	(1.581)	(1.650)	(1.285)	(1.781)
Month	Included	Included	Included	Included
Time_of_day	Included	Included	Included	Included
Weekend	Included	Included	Included	Included
Observations	3,037	3,037	3,037	3,037
LR Chi2	162.77***	189.82***	199.72***	213.88***
Log-likelihood	-2246.573	-2230.506	-2225.228	-2216.394

BIOGRAPHICAL SKETCH

Massara Alazazi has earned her Doctor of Philosophy degree in Business Administration with a focus on Information Systems from the University of Texas Rio Grande Valley in May 2023. She received both a Master of Science degree with honors in Information Systems in 2014 and a Bachelor of Science degree in Computer Science in 2012 from Dakota State University.

Massara has authored and co-authored several papers that appeared in the proceedings of major IS conferences including ICIS, HICSS, AMCIS, and DESRIST. Her research interests focus on social media and big data analytics, health IT and online health communities, online crowdfunding, and cloud computing.

Massara worked as a research assistant and a teaching assistant during her doctoral study. She received several awards for her outstanding research and teaching including Ph.D. Student Research Award, Spring 2022, Ph.D. Student Research Award, Fall 2021, and Ph.D. Student Teaching Award, 2021.

Outside of academics, Massara enjoys outside activities with family and friends including walking, swimming, and traveling.

Email address: musirra@hotmail.com