MODELING TWITTER DATA FOR EFFECTIVE DISSEMINATION OF HEALTH-RELATED INFORMATION

by

Lida Safarnejad

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Computing and Information Systems

Charlotte

2020

Approved by:

____

Dr. Yaorong Ge

____

Dr. Shi Chen

____

Dr. Weichao Wang

____

Dr. Wlodek Zadrozny

ABSTRACT

LIDA SAFARNEJAD. Modeling Twitter Data for Effective Dissemination of Health-Related Information. (Under the direction of DR. YAORONG GE and DR. SHI CHEN)

Social media platforms, such as Twitter, are attracting a growing number of people with diverse demographic characteristics to share and obtain information about various topics. As a result, these platforms have become one of the main targets for practitioners and decision-makers from various fields, such as politics and public health, to study public opinion and at the same time to spread their messages. Two challenges in utilizing social platforms as a means of communication are how to craft and how to deliver a message such that it reaches a great number of audiences and keeps them engaged. Addressing these problems is hardly possible without thoroughly analyzing how a piece of information goes viral on a social platform. This doctoral dissertation aims to model the dissemination of health-related information on Twitter from various perspectives. First, I investigate driving forces of the general public's engagement on social media during health emergencies. The two contributors that I consider are 1) real-world events, such as announcements by World Health Organization, and 2) the role of highly active users and also those who receive great attention from other users. Second, I systematically model and investigate information cascades through the retweeting processes of tweets. My analysis in this part reveals that the propagation patterns of tweets carrying misinformation are different from those containing true information. I propose a framework to operationalize and test the hypothesis "misinformation tweets are propagated differently from true information tweets." Finally, based on the differences that I discover by analyzing (mis)information propagation, I propose a feature-rich machine learning model to identify misinformation on Twitter. The above three perspectives offer a holistic overview of the main challenges and prospective/feasible solutions for beneficially

applying social media in public health.

ACKNOWLEDGEMENTS

I would like to thank the following people, without whom I would not have been able to complete my doctoral research and completing my Ph.D. degree. First, I would like to express my sincere gratitude to Dr. Yaorong Ge, my advisor, for his continual support, encouragement, and patience. Moreover, I am incredibly grateful to Dr. Shi Chen for his dedicated support and guidance. I would like to thank my other doctoral committee members, Dr. Wlodek Zadrozny and Dr. Weichao Wang, for their feedback and contribution. I would also like to say a special thank you to Dr. Mary Lou Maher, the SIS department chair, for her support and wisdom. I would also like to thanks Dr. Qian Xu, Dr. Siddharth Krishnan, and Dr. Arunkumar Bagavathi for their valuable contribution to my work. I also wish to thank my labmates at Health Informatic Lab and my friends at UNC Charlotte, especially Maryam, Chuqin, Tiyani, Yi, Jody, Jessica, Jingyi, Yusra, Malak, Yasaman, Hossein, and Haadi, without whom this long journey would have been even harder.

I would like to thank my beloved parents, Jalil and Forough, for their unconditional love and support and also my wonderful sister and brother, Ladan and Saman. Finally, I cannot forget to thank my lovely husband, Amirreza, for his support, encouragement, and patience during these stressful years.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

API an acronym for Application Programming Interface

AUC an acronym for Area Under Curve

BET an acronym for Betweenness Centrality

CA an acronym for Content Analyzer

CCF an acronym for Cross-Correlation Function

CDC an acronym for Centers for Disease Control and Prevention

CDF an acronym for Cumulative Distribution Function

DEN an acronym for Network Density

DIA an acronym for Network Diameter

DSI an acronym for Data Science Initiative

EBM an acronym for Evidence-based Medicine

FN an acronym for False Negative

FP an acronym for False Positive

FPR an acronym for False Positive Rate

FWHM an acronym for Full Width at Half Maximum

GBS an acronym for Guillain Barre Syndrome

GM an acronym for Genetically-Modified

GMO an acronym for Genetically-Modified Organism

HPP an acronym for Homogeneous Poisson Process

K-S an acronym for Kolmogorov-Smirnov

LDA an acronym for Latent Dirichlet allocation

LIWC an acronym for Linguistic Inquiry and Word Count

MI an acronym for Mutual Information

MOD an acronym for Network Modularity

MST an acronym for Minimum Spanning Tree

NHPP an acronym for non-homogeneous Poisson proce

NIF an acronym for Network Influence

NLP an acronym for Natural language processing

NN an acronym for Neural Network

OUT an acronym for Out-degree Centrality

PD an acronym for Peak Detector

PHEIC an acronym for Public Health Emergency of International Concern

PMI an acronym for Pointwise Mutual Information

POS an acronym for Part-of-Speach

REA an acronym for Network Reach

RF an acronym for Random Forest

RNN an acronym for recurrent neural network

ROC an acronym for Receiver Operating Characteristic

RT an acronym for Retweet

SC an acronym for Signal Constructor

SVM an acronym for support-vector machine

TM an acronym for Top Mentioned

TN an acronym for True Negative

TP an acronym for True Positive

TPR an acronym for True Positive Rate

TR an acronym for Top Retweeter

TRRT an acronym for Top Received Retweets

TT an acronym for Top Tweeter

VIR an acronym for Structural Virality

VIS an acronym for Visualizer

WHO an acronym for World Health Organization

WIE an acronym for Wiener Index

WT an acronym for Wavelet Transform

ZIKV an acronym for Zika Virus

CHAPTER 1: Introduction

A large population of people from a wide demographic is relying on social media to obtain information regarding various topics and also share their opinions. At the same time, practitioners and decision-makers from various fields are considering social media as one of the main conduits to release news and communicate with their audiences.

One of the interesting aspects of using social media to disseminate information is the high velocity at which a piece of news can get viral and reach to the large number of users who are mainly dependent on these platforms to seek and obtain information. Nevertheless, this property of social media has a serious downside; misinformation can also get viral at an uncontrollable speed and impact a great number of users. In [1], Lazer *et al.* define misinformation as false or misleading information which is fabricated such that they are indistinguishable from legitimate information for users who have a limited, if none, knowledge in a specific field. User-generated information in these platforms is disseminated without going through any fact-checking process [2]. This property makes social platforms susceptible to misinformation or fake information dissemination. In other words, malicious actors exploit this property to propagate misinformation. Moreover, due to the lack of knowledge, some benign users might believe misinformation and unintentionally help its propagation by re-sharing it [1]. Misinformation propagation has been a major concern among practitioners in various fields. To illustrate, politicians are using social platforms such as Facebook and Twitter to crush their competitors and manipulate public opinion. In [3], Grinerg *et al.* demonstrate that during the US presidential election in 2016, 0.01% of Twitter users shared almost 80% of fake news, and they could engage a great population of eligible voters. Another field that misinformation propagation could even jeopardize people's lives and create a disaster is public health. To illustrate, one of the greatest human achievements in fighting with and controlling disease outbreak is vaccination [4]. In [5], Broniatowski *et al.* look into the role of Russian trolls in propagating fake news which cast doubt on and support the debate against vaccination. In another study [6], Sharma *et al.* show that during ZIKV outbreak in 2016, the videos related to Zika contained more fake information rather than true and informative. These are a few examples showing how much fighting back with misinformation propagation is vital. To this end, practitioners and decision-makers must carefully and actively monitor social platforms to intervene and neutralize the impact of misleading information by feeding users with trustworthy and useful information. In other words, scholars argue that using social media is a double-edged sword: although malicious actors use social media platforms to propagate misinformation, they can be used as a tool to contain and prevent misinformation as well [7], [6]. Right and relevant information must be crafted carefully and released at the right time to pick the public's attention and influence their viewpoints. Another key factor that should not be ignored is the role of users in relaying information. In the real-world, opinion leadership is determined by social values; however, in social platforms, equations are different. It is highly common that individuals with no considerable fame in the real world can be quite influential in the context of social media. On the one hand, misinformation can be contained by identifying users who trigger and promote misinformation in social media. On the other hand, influential users can help lessen the effect of misinformation by providing and facilitating the propagation of true information.

One domain that can highly benefit from using social platforms to study public opinion and also conveying messages to their audiences is public health. Considering the large population and the wide demographic of social media users, social media platforms are invaluable resources for public health professionals to study public opinion regarding health issues and also to better communicate with the public, especially during health crises such as the Zika outbreak in 2016 or the current COVID-19 pandemic. Scholars in this field have realized and are emphasizing the crucial role of using social platforms by health organizations to communicate with people to boost the public's knowledge of health [8], [9]. Health practitioners have gone even further and used social platforms, such as Facebook or Twitter to characterize users who might develop a kind of disease [10], [11] and also predict disease outbreaks [12], [13]. This doctoral dissertation spans over designing and developing computational models to effectively employ social media in the public health domain.

First, I investigate the driving factors that have a considerable impact on health-related discussions on social media. The three main factors that I consider are 1) the specific health issue by itself, such as the Zika outbreak or the current COVID-19 pandemic, 2) real-world events, such as a Public Health Emergency of International Concern (PHEIC) announced by WHO, and 3) users who play key roles in heating the discussions.

Second, I propose an algorithm to model information cascades on social media. By analyzing the information cascades, then I demonstrate how misinformation is propagated differently from true and relevant information.

Third, I propose a new set of features extracted from information cascades. Then, to show the effectiveness of the proposed features, I build a misinformation detection system that can predict misinformation with high accuracy.

Without loss of generality, I narrow down my concentration to Twitter, which includes a wide range of demographics of people, as pointed out by researchers such as [14]. This micro-blogging service has 100 million active users posting 500 million tweets every day [15]. Thus, it has become one of the main targets of practitioners and decision-makers to study public opinion and foster public relations.

To practically demonstrate the effectiveness of my proposed frameworks in analyzing the Twitter platform, I analyze Zika-related discussions on Twitter in 2016 during Zika Virus (ZIKV) outbreak.

In the rest of this chapter, I review existing state-of-the-art research works closely related to the research questions that I plan to address in my dissertation.

1.1 Trend Detection on Twitter

Proposed methods and frameworks in the area of event detection, generally rely on signal processing, Natural Language Processing (NLP), and text mining techniques to analyze social media discussions [16], [17]. In this section, I briefly review some of these research works.

Mathioudakis *et al.* , in [18], proposed TwitterMonitor, which is a system to detect trends on Twitter in real-time. TwitterMonitor achieves this goal by first identifying bursty keywords, the ones that appeared much more often in the tweet stream in recent history. Then, it groups such keywords based on the number of their co-occurrence in tweet stream to form trending topics.

In [19], Weng *et al.* model words occurrence in tweet streams using signals to capture the temporal changes in their appearance. Afterward, they use cross-correlation to find similarity between word-specific signals and therefore determine events. In a close work [20], Cordeiro *et al.* consider bursts in hashtags occurrence, specified by wavelet signals, to be events. Next, Latent Dirichlet Allocation (LDA) topic modeling is performed to uncover the underlying topic of the identified events. In [21], a pipeline is proposed to detect and visualize events. They construct a time series from texts related to a topic within a specific time window. Then, the cumulative sum control chart is utilized to spot temporal changes in the topic-specific signals. With a purely NLP approach, Ren *et al.* in [22] propose an LDA based topic modeling and adopt Support Vector Machine for Twitter sentiment classification. In a close work, authors in [23] cluster tweets using the Locality Sensitive Hashing technique.

1.2 Influential Users

In the real world, opinion leaders are determined based on their social values; however, in social platforms, criteria based on which opinion leaders are identified are different. To emphasize the difference between them, the former opinion leadership is called offline authority, while the former one is called online authority [24]. To be more specific, opinion leadership in social media is determined by an individual's ability to influence on spreading information [24]. As an example, Cha *et al.* explains that adhering to a single topic is important. Their results also show that when a tweet receives a great number of retweets, it is mainly due to its textual content; however, when someone receives high mention, it is more of the identity of that node. In social platforms, leadership is more of credibility rather than authority [25]. Disseminating true and on time information is two factors that give an actor credibility. In addition, the level of persistency and activity of a user is also important. In [26] Xu *et al.* explains that how much an individual is active on a social platform, how many connections she has, and how much useful are information that she shares are the factors that let someone acts as an opinion leader. To measure the level of credibility of an actor, I can look into the number of retweets his messages receive.

Authors in [27] explains that opinion leaders not only are successful in getting attention but also can redirect others' attention and influence the action that they take by providing recommendations. Ordinary users may not have access to expensive or specific resources (e.g., a TV channel) to share their opinion, but social platforms have provided such users with a fast and cheap tool to share their viewpoints. Therefore, the degree of being a leader is directly related to the usefulness and credibility of information that one provides. Park *et al.* also believe that how much a user is active influences her degree of being influential. They also propose the idea of multi-step flow as opposed to two-step flow [28]. Multi-step flow means several users with a good number of connections play a crucial role in relay information. Their analysis shows that demographic features, such as age and gender, do not have that much to do with leadership; however, the number of followers and the level of activity of someone in posting and sharing tweets are positively associated with being a leader on Twitter. Vergeer explains that users with more connections tend to attract even more connections, or in other words, the rich get richer. Moreover, offline and online authority are not separated, and they can influence each other [29].

In [30], Choi *et al.* explains that one can identify opinion leaders by investigating how information flow. Opinion leaders on social media are not merely determined by their social characteristics and values, but by some network-based attributes on social platforms. They define an opinion leader as a node that facilitates the flow of information. Their study shows that betweenness centrality has a positive association with being opinion leader because the flow of information in the directed network of mention that they constructed was not possible unless that node (opinion leader) was present. The in-degree centrality was also important to identify opinion leaders; people whose messages are frequently retweeted are also candidates to be opinion leaders. However, there is not that much difference between the content that they create with others.

Opinion leaders in social media do not necessarily have a high social status. If we identify opinion leaders, we can also predict users' involvement. When someone has more followers, there is a great chance that she gets more attention because more people will see her tweets [31].

In [32], Karlsen *et al.* explains that users in social platforms are exposed to the news even if they do not want to because they see what people who they are connected to share. Opinion leaders on social media are those who are highly active.

In [33], they create a network of emergency physicians and their followers. They consider several factors to determine who is influential: Eigenvalue, betweenness, and in-degree centrality. Their results show that one single metric is not a good criterion to measure the influence of a node, but we should consider them all. In [34], Heidemann *et al.* say users' degree of influence is determined based on their connectivity and level of activity. They use the idea of PageRank to identify opinion leaders.

In [35], authors proposed OLFinder. They first perform topic modelings, next find users who post in each topic and categorize users, and then they calculate a competency score for each user. Then they calculate in-degree centrality. They calculate a leadership score based on these two features.

It is noteworthy to mention that Opinion leaders are not necessarily experts in a specific domain. In [36], authors define peer-experts and propose a technique to identify them in health-related forums. They define peer-experts as users who are not officially trained but have good knowledge in a particular field. To this end, they rely on the features related to the text, such as its semantic, and keywords. They look at other features, such as the number of followers. They build a friendship network and calculate PageRank for users. They also build a monthly time series from the number of posts by a user. From the time series, they extract features such as mean and skewness. Considering all user, text, and temporal related features, they conduct a supervised classification to identify peer-experts from novice users.

Similar to the concept of peer-experts in the field of health, Okazaki *et al.* in [37] talk about the role of prosumers as leadership to engage people in the field of marketing. These users can get others' attention by their recommendation. They also emphasize that prosumers are not opinion leaders but are customers who are indirectly collaborating with a company by providing their positive feedback and recommendation.

1.3 Information Dissemination on Twitter

Researchers have tried to formulate and address the problem of misinformation dissemination from various aspects. Authors in [38] propose an automatic method to verify the credibility of tweets on Twitter. They use TwitterMonitor [18] to detect trends and then using both user-related features, such as the number of followers, and text-related features, such as URLs within the message, to classify detected trends into credible or non-credible. Odonovan *et al.* in [39] demonstrates that URLs, mentions, retweet, and tweet length are features that could be used to determine the credibility of a tweet. Allcot *et al.* , in [40], conducted a study to investigate the volume of misinformation circulated on social media, Facebook and Twitter, between 2015 to 2018. To this end, they curated a list of 570 fake news sites and considered tweets/posts containing a link to these websites as misinformation. Their analysis shows that the interaction these fake news websites have started to increase in early 2015, and this trend continued until one month after the election in 2016.

In [41], Gupta *et al.* hypothesize that rumors and misinformation increase when a major event occurs. To validate their hypothesis, they gathered tweets posted about particular events and propose a system, called TweetCred, to assign a credibility score to the collected tweets. Using SVM and based on a set of features, TweetCred assigns a credibility score to each tweet. RumorLens [42] is a pipeline that combines human efforts and automatic computational techniques to classify and visualize rumors propagated on Twitter. Twitter-Trails [43] is a tool to detect misinformation on Twitter. It does not directly classify tweets into misinformation or benign but provides a tool to gather information upon a trend; such information then can be consumed by an analyst to judge how much credible a tweet is. Some other similar systems that have been proposed to detect misinformation and fake news are FactWatcher [44] and Hoaxy [45].

In [46], Shao *et al.* investigate the role of bots in propagating misinformation. They analyze tweets that point to low-credible articles. They identify such articles by crawling low-credible websites and then use Hoaxy [45] to gather information about the propagation of that information. They find tweets that are linked to these articles and investigate accounts that post those tweets. They assign automation score to each account to identify bots. In [47], researchers constructed a recurrent neural network (RNN) to classify rumor versus real tweets.

In [48], authors explain that bots, accounts, which are managed by software, exhibit behaviors that distinguish them from human accounts. Accounts are classified based on a set of user-based, network-based, temporal, and textual-based features.

In [49], a technique is proposed to detect controversial trending topics or potential rumors. By controversial, they mean people express doubt about the truth of a post. Using regular expression, they identify clusters of tweets that are casting doubt on a particular topic and clusters of rumors. The clusters, then, are expanded by adding other tweets that are related to the topic. These clusters of rumors are ranked based on a set of features. These controversial topics are then examined by analysts to confirm whether they are truly rumors or not.

In [50], Wu *et al.* create networks from tweets, find embedding for nodes. These node embeddings, along with other features, such as textual content of tweets and posting times, are then used to classify messages to fake and non-fake news.

A close work to ours is that of Vosoughi *et al.* [51], which track cascades of some known rumors. They define cascade as an unbroken chain of retweets of a tweet. They define some measurements, such as depth and size, to quantify and compare cascades. Their analysis shows that false information spread faster and farther [51]. Their work is different from ours in the sense that I aim to model (mis)information propagation using networks and graphs and then investigate the propagation of a piece of information by studying characteristics of such networks.

1.4 Dissertation Overview

In this doctoral dissertation, I aim to analyze health-related information dissemination on Twitter mainly from three aspects: 1) investigating driving factors that create discussions and drive user engagements with health-related topics, 2) modeling information diffusion on Twitter to compare and contrast health-related misinformation versus true information propagation, 3) design and develop a misinformation detection system which can identify health-related misinformation with high accuracy.

In chapter 2, I propose and examine two hypotheses about driving factors that impact discussions on Twitter. The two forces that I investigate are 1) real-world events, such as critical announcements released by health organizations, 2) users whose tweeting/retweeting activities could lead to further engagements of other users by health-related discussions.

In chapter 3, I propose a solution to model information dissemination on Twitter. To be more specific, I present a method to estimate information diffusion networks. By systematically analyzing the structures of those networks, I reveal that diffusion patterns of misinformation are different from true information.

In chapter 4, I propose a set of salient features extracted from information diffusion patterns to build misinformation classification models. In specific, I propose a method to model the retweeting process of a tweet as a Non-Homogeneous Poisson Process (NHPP). A subset of features that I propose in this chapter is extracted from these NHPPs. To evaluate the presented features, I build several classifiers based on the newly proposed and existing ones.

CHAPTER 2: Identifying Influential Factors in the Discussion Dynamics of Emerging Health Issues on Social Media: Computational Study

2.1 Overview

This chapter aims to provide a methodological framework and insights to better understand the driving forces of web-based public discourse during health emergencies. Therefore, health agencies could deliver more effective and efficient web-based communications in emerging crises. In total, two hypothetical drivers were proposed and examined: 1) sporadic but critical real-world events, such as the 2016 Rio Olympics and World Health Organization's Public Health Emergency of International Concern (PHEIC) announcement, and 2) a few influential users' tweeting activities. I used the Zika virus epidemic in 2016 as a case study to formulate and test the proposed hypotheses.

2.2 Introduction

2.2.1 Background

Social media platforms, such as Twitter and Facebook, are attracting a growing number of people with diverse demographic characteristics to share and obtain information on the web. As a result, these platforms have become one of the main targets for practitioners and decision-makers across various fields to understand public opinion and, at the same time, disseminate information to the public [52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68]. Many public health agencies and organizations, such as the US Centers for Disease Control and Prevention (CDC), are active on Twitter and other social media platforms as the main channels of communication with the general public, especially during health emergencies such as the 2014 Ebola and 2016 Zika outbreaks. The CDC has 67 officially associated Twitter accounts that cover a wide variety of health- and disease-related topics [69]. In February 2016, when Zika caused 5168 confirmed noncongenital cases in 50 states and the District of Columbia in the United States, and a much higher case number across US territories [70], Dr. Tom Frieden, the former CDC director, hosted live Twitter chats with the general public regarding Zika [69].

Nevertheless, there are multiple challenges in utilizing social media platforms as an effective channel of communication. A considerable percentage of users are unfamiliar with the emerging health issue. At the same time, user-posted content does not go through any rigorous fact-checking process, making room for misinformation to take advantage of such a situation. During the 2016 Zika epidemic, despite the CDC's outstanding online presence, inaccurate information about Zika proliferated on social media and outperformed CDC (and other official sources such as the World Health Organization (WHO)) by a large margin [69]. Uncertainty about the root cause and transmission route of this virus gave room for the proliferation of rumors and misinformation [71, 72].

In addition to the problem of misinformation propagation, the rhetorical aspect of a message, or in other words, crafting it based on the needs and perception of audiences is, a critical challenge [73, 74]. Studies have shown a significant topic discrepancy between public concern about Zika and responses by CDC on Twitter [17, 68, 69]. More specifically, the general public was more concerned about the transmission routes of Zika and effective prevention methods, whereas the CDC focused on symptoms to educate the public [75, 76]. Glowacki et al. [76] argued that this could be seen as the failure of the CDC to identify what kind of information the public was looking for and respond accordingly or it could be an on-purpose attempt by the CDC to redirect public attention to what the CDC believed to be more important during the epidemic.

In addition, one important yet overlooked issue in utilizing social media platforms as a communication mechanism with the public is the low rate of user engagement while social media should be interactive and engaging environments for the public interaction rather than being one-directional news outlets [77, 74, 60, 78, 79]. To better engage the public, it is essential to recognize critical factors that are directing and driving the general discussion dynamics on social media. Such factors can be discovered by observing and analyzing the public's tweeting behaviors on social media [61, 80, 78]. Learning about these factors can help health agencies to accurately predict shifts in the public's concern about health issues and provide the public with useful information accordingly. As a result, systematically collecting and analyzing data related to the public discourse of emerging health issues on social media, also referred to as digital public health surveillance [62], is essential for understanding public concerns and disseminating useful information correspondingly.

2.2.2 Objectives

In this chapter, I aimed to identify important factors that could potentially drive tweeting dynamics in the 2016 Zika epidemic. I comprehensively analyzed all Zika-related English tweets posted during 2016. I further proposed and evaluated the following two testable hypotheses (H):

1. H1: The tweeting dynamics of Zika was associated with and influenced by a few real-world critical events, other than the continuous Zika epidemic.
2. H2: The tweeting dynamics of Zika were driven by a few highly influential users (colloquially referred to as influentials hereafter), which led to the public discourse of Zika on Twitter.

2.3 Data Acquisition

More than six million English tweets, including the keyword Zika from January 1 to December 31, 2016, were retrieved via the Gnip application programming interface

(API) through the Data Science Initiative (DSI), University of North Carolina Charlotte. All associated metadata with these tweets, such as retweet counts, posted time, and the verification status of tweeting/retweeting ID, were also included in the data set. This data set represented the complete public discourse about Zika in English and was therefore not as prone to potential selection bias as the common 10% sample provided by the common Twitter API. Therefore, the data set in this study was able to provide an unbiased and comprehensive depiction of the public's discourse of Zika, the most discussed health topic in 2016 on a major social media platform.

2.4 Association Between Critical Events and Tweeting Dynamics

Health emergencies, such as the Zika epidemic, would never occur in isolation and almost always be intermingled with other health, social, societal, and political events in the real world. I suggest that related and sometimes unrelated real-world events could be potential driving forces of Zika discussions on social media. Unlike the time series of daily Zika case counts, these real-world events were much more discrete and sporadic. Here, I evaluated the first hypothesis (H1) such that Zika-related tweeting activities were substantially influenced by sporadic real-world events. I adopted the definition of an event provided by Hasan et al. [17] stating, "An event, in the context of social media, is an occurrence of interest in the real world which instigates a discussion on the event-associated topic by various social media users, either soon after the occurrence or, sometimes, in anticipation of it."

2.4.1 EventPeriscope Pipeline

I developed an analytical pipeline, EventPeriscope, to explore and quantify the impact of real-world events on the tweeting dynamics of a specific topic (e.g., Zika in this study) and to evaluate H1. Fig. 2.1 shows the main components of EventPeriscope: Signal Constructor (SC), Peak Detector (PD), Content Analyzer (CA), and Visualizer (VIS). To capture temporal dynamics of discussions of a specific topic

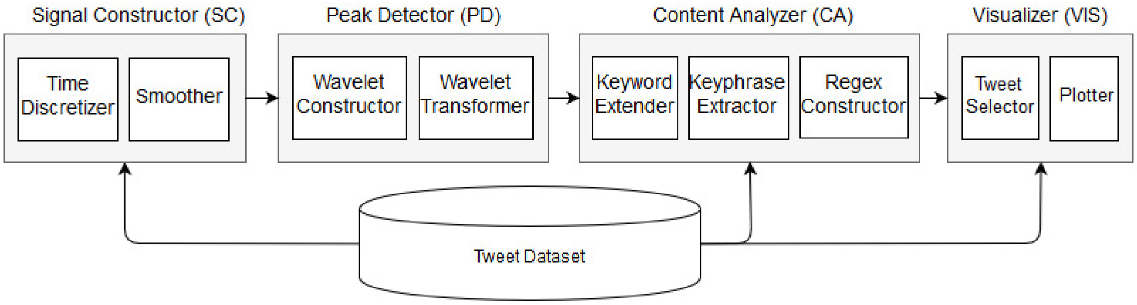


Figure 2.1: EventPeriscope Pipeline.

on Twitter, SC module constructs a signal from the tweets posted in a particular time window specified by the user. The constructed signal is then passed as an input to the PD module. This module uses the wavelet transform to locate peaks in the input signal. Observing a peak in close proximity to the time point when the hypothesized real-world event happens is necessary but insufficient to conclude that the event has directly caused the discussions. CA module examines tweets posted within a short time interval, specified by the user, around the time when the event of interest occurred to create a set of regular expression (regex) rules to detect tweets discussing the event. Afterward, all tweets in the tweet stream dataset are tested against the regex rules to find all matched tweets, and subsequently, estimate the percentage of tweets related to this event. In the following sections, each module is explained in detail.

2.4.1.1 Signal Constructor

To construct a signal from a tweet stream, first, the time interval between the first tweet and the last tweet in the stream is partitioned into fixed-length time slots or bins, and a signal is created from tweet counts in each bin. We use two attributes, magnitude and width, to characterize a peak in the tweet signal. In other words, we are interested in rises that are above a certain threshold in the signal and also persist for more than a pre-specified time interval. Therefore, the signal is smoothed to filter out small oscillations that last for a short period of time. For Smoothing, I employ Kernel Density Estimation (KDE) [81]. KDE is a smoothing technique that estimates the value of each data point by the average or weighted sum of its value and the values of its neighboring data points. For every data point, its new value $\hat{f}(x)$ is calculated by

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{x - x_i}{h}), \tag{2.1}$$

where $K(.)$ is a kernel function, and h is the smoothing parameter. h determines smoothness of a curve. Based on the shape of the peaks observed in the constructed tweet signal, I use Gaussian kernel for smoothing:

$$K(U) = \frac{1}{\sqrt{2\pi}} \exp(u^2), \tag{2.2}$$

The resulting smoothed signal is then passed to PD module to locate peaks.

2.4.1.2 Peak Detector

PD module utilizes the wavelet transform to detect peaks in the signal constructed by SC. In the wavelet transform, wavelet coefficients associated with a function $f(x)$, the tweet signal in my case, are calculated by

$$W_f(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(x)\overline{\Psi(\frac{x - b}{a})}dx, \tag{2.3}$$

where $\overline{\Psi(t)}$ denotes the complex conjugate of the mother wavelet $\Psi(t)$, a is the scale, and b is the time shift. Coefficients at different scales and time shifts are represented by a matrix. When a peak appears in the signal, maximum values of different scales occur at close time shifts, resembling a ridge in the coefficient matrix [82]. In this paper, I adopt Mexican hat wavelet as the mother wavelet:

$$\Psi(t) = \frac{2}{\sqrt{3\sigma}\pi^{\frac{1}{4}}}(1 - (\frac{t}{\sigma})^2) \exp(\frac{-t^2}{2\sigma^2}), \tag{2.4}$$

where σ is the scale.

2.4.1.3 Content Analyzer

CA module interactively constructs a set of regular expression (regex) rules describing an event. Subsequently, it utilizes the constructed rules to retrieves all tweets in the stream that are related to the event of interest.

CA module requires two user-provided inputs: 1) a set of keywords, which can be single words, hashtags, or mentions, related to the real-world event of interest, and 2) a time interval around the event that CA module uses for its analysis. Keyword Extender in CA module extends the input keyword set by examining tweets posted within the specified time interval to find other keywords that have the highest correlation with the user-specified keywords. To this end, Keyword Extractor uses Pointwise Mutual Information (PMI) to quantify the association between keywords. PMI is a statistical approach for measuring the level of dependency between two observations [83], in my case two words. PMI between two words w_1 and w_2 is calculated by:

$$PMI\left(w_1,\ w_2\right) = log\ \left(\frac{p\left(w_1, w_2\right)}{p\left(w_1\right) \times \left(w_2\right)}\right), \tag{2.5}$$

where $p(.)$ is probability function. Two words that have a high PMI value are strongly associated with each other. In other words, they co-occur frequently. Keywords are sorted based on their PMI value, and a set of keywords with high PMI value are selected. Keyphrase Extractor extracts keyphrases containing the keywords generated by Keyword Extractor. In this step, I adopt the NLP technique [84]. First, the textual contents of tweets are lemmatized and tagged with a Part-of-Speach (POS) tagger. Then, Keyphrase Extractor utilizes a predefined context-free grammar [84], to retrieve all noun phrases. Next, noun phrases containing the event-relevant keywords are extracted and sorted based on their frequencies. The user will be provided with this list to choose keyphrases that describe the event. Matching all tweets in the dataset with selected keyphrases to find relevant tweets is technically challenging as such keyphrases can be rewritten in multiple forms, and many of the possible forms may not be presented in the short time interval that is considered to develop event-relevant keyphrases. For instance, spaces between words in a keyphrase might be dropped (e.g., OlympicGames), spaces might be replaced by other characters such as dot or dash (e.g., Olympic.Games), or other words might be inserted between the words (e.g., Olympic Summer Games). To handle such variation and combinations, Regex Generator constructs a set of regular expressions (regex) rules based on the selected keyphrases. Moreover, the user can customize the generated rules.

2.4.1.4 Visualizer

VIS module utilizes the resulted regex rules to capture tweets relevant to the event of interest. To be more specific, VIS tests all tweets in the dataset against these regex rules, and constructs time series from the number and percentage of matched tweets. The constructed event-specific time series demonstrate when the event appeared in the discussions and what percentage of discussions were dedicated to this event. Moreover, peaks in these time series indicate the time points when the event had elevated discussions. Analysts can then carry out more investigation to uncover the underlying reason for the observed rises.

2.4.2 Case Studies of Critical Events

Real-world events could be categorized into 2 dichotomized and mutually exclusive types [16]: (1) planned (ad hoc) events that people expected in advance, such as the 2016 Rio Olympics; (2) unplanned (posthoc) events that people would not know beforehand, contrary to planned events. An example of unplanned events was the WHO's Public Health Emergency of International Concern (PHEIC) announcement about Zika on February 1, 2016. In the next section, I have discussed methodological differences in exploring planned (Rio Olympics 2016) and unplanned (WHO-PHEIC) events in detail. Planned events might increase their presence in tweeting around the event, but it could be mentioned throughout the entire year because people were well aware of it beforehand. Unplanned events, however, would not be mentioned in tweets until their occurrence in the real world. In the next section, I examine the impact of these two types of real-world events on Twitter discussion dynamics.

2.4.2.1 Unplanned Event: WHO-PHEIC Announcement

On Feb. 1, 2016, the Director-general of WHO, Margaret Chan, declared a PHEIC of potential Zika pandemic [85]. In this statement, in addition to raising concerns over the linkage of Zika with microcephaly and other neurological disorders, WHO provided travel advice in Zika-impacted regions. The WHO PHEIC announcement was an unplanned event, and the general public did not have prior knowledge of its occurrence. Therefore, it could have mainly influenced tweets posted after the PHEIC announcement. I used EventPeriscope to quantify the influence of the WHO-PHEIC event on Twitter Zika discussions; the details are described as follows.

First, a signal was constructed from posted Zika-related tweets, which is referred to as the *main tweet signal* hereafter. The main tweet signal peaked almost immediately after WHO-PHEIC event on day 32 (Feb. 1, 2016), indicating a potential correlation between the event and Zika discussions. The textual contents of tweets were then analyzed to verify the association between Zika and WHO-PHEIC. The set of tweets posted in a two-day interval, the day of WHO-PHEIC announcement (Feb. 1) and one day after (Feb. 2), were used as the input of the CA module to construct a regex rule describing the WHO-PHEIC event. In addition, this module was given a set of two keywords, *WHO* and *PHEIC*, relevant to the WHO-PHEIC announcement (event). To find other relevant keywords, the Keyword Extractor in the CA module used pointwise mutual information (PMI), and calculated PMI values between each of these two keywords and all the keywords extracted from the input tweet set. The keywords were then sorted from largest to smallest based on PMI values, and the ones with the highest PMI values were selected.

The additional set of keywords found using the above approach included emergency, public, international, global, world, and health. A single word within a text is not usually descriptive enough to reveal the main topic of the text. Therefore, to take the context of a tweet into consideration, and obtain a more accurate result, the Keyphrase Extractor uses the keywords to extract keyphrases describing the event. We define a keyphrase to be a noun phrase which contains at least one of the keywords. Obtained Keyphrases relevant to WHO-PHEIC are public health emergency, global emergency, and world health. Based on these key-phrases, a regex rule was crafted. By using a similar approach, another regex rule was generated to capture tweets that were talking about WHO (regardless whether it was related to WHO-PHEIC or not). Finally, the VIS module tested all tweets in the tweet stream dataset with the resulted regex rule, and created two daily time series; one demonstrating the dynamics of tweets about the WHO-PHEIC and the other one pertaining to WHO.

2.4.2.2 Planned Event: RIO2016

The Rio 2016 Olympic Games were held from Aug. 5- Aug. 21, 2016 in Rio de Janeiro, Brazil, amidst global concerns about Zika outbreak. In November 2015, Brazilian authorities declared a national public health emergency due to a high rate of confirmed Zika cases [77]. As RIO2016 was a planned event, I expected to see tweeting about Zika and RIO2016 before its opening. The CA module of the EventPeriscope pipeline was initialized with tweets posted from Aug. 4 to 6 (days 217 to 219) within plus or minus a 1-day window of the RIO2016 opening. Then, a regex rule was generated to detect the co-occurrence of the Zika and Rio Olympics topics in Twitter discussions. The final keywords and key phrases were Rio, Olympics, Rio2016, 2016 Olympics, and Rio Olympics.

2.5 Association between Online Influentials and Dynamics of Zika Discussion on Twitter

In this section, I hypothesized (H2) that a few influentials on Twitter made a substantial contribution in driving the tweeting dynamics, that is, a noticeable sudden rise in the number of tweets. To evaluate this hypothesis, I defined 4 different types of web-based influentials in 2 major categories: active influentials who posted a large number of original tweets about Zika (top tweeter [TT]) and who retweeted a lot about Zika from other accounts' posts (top retweeter [TR]). These users actively disseminated Zika-related information to the public. In addition, influentials on social media could be passive as well: whose original posts were retweeted a lot (top received retweets [TRRT]) and who received many mentions (@_userID) from other Twitter users (top mentioned [TM]). These passive influentials, on the other hand, were more reflective of public perception and engagement of Zika discussions on Twitter. I ranked and selected top 100 users in each of these four influential groups. The tweeting dynamics of each user in the TT, TRRT, and TM group and the retweeting dynamics of each user in the TR group were then extracted and examined. These tweeting/retweeting dynamics were then aggregated and compared with the overall tweeting dynamics using cross-correlation function (CCF) in each quarter of 2016 as well as the entire year.

CCF measures the temporal similarity between the two time series, as shown in equation 2.6. To calculate $CCF_{X,Y}(k)$ between two discrete time series $X[t]$ and $Y[t]$, first the time series $Y[t]$ is shifted k units, where $-\infty < k < \infty$, to the left in time, and then the correlation between $X[t]$ and $Y[t+k]$ is computed. $CCF_{X,Y}(k)$ at time shift k is calculated by

$$CCF_{X,Y}(k) = \sum_{t=-\infty}^{+\infty} X^{*}[t]Y[t+k], \tag{2.6}$$

where $*$ represents the complex conjugate of a function. Positive correlation value shows that the signal which is shifted in the calculation of CCF is ahead of the other signal. Maximum cross correlation occurs in the time lag k where $X[t]$ and $Y[t+k]$ are similar the most.

I defined $Y_{g,i}(t)$ to be the time series of i-th user in group g, where $g \in \{TRRT, TM, TT, TR\}$ and then calculated $CCF(X(t), Y_{(g,i)}(t))$, where $CCF(.)$ denotes cross correlation function. This let us to derive the time lag between each user's tweeting (or retweeting) dynamics and the main tweet signal to test whether these influentials' tweeting activity preceded the overall tweeting dynamics. This step was critical to further reveal if these influentials actually initiated an increasing number of Zika-related tweets, or the other way around, that is, these influentials were actually following and catching up with the general trend on Twitter.

Moreover, to test the group-level association between influential type and the overall tweeting dynamics, or in other words, to examine how each group of users are acting as a whole, for each group, I also calculated cross correlation between $Y_{g,total}(t)$ and $X(t)$, where $Y_{g,total}(t)$ is the aggregation of all $Y_i(t)$ and calculated by:

$$Y_{g,total}(t) = \sum_{i=1}^{100} Y_i(t)$$

I also examined the overlap between the four groups of influentials by calculating the intersection of any two sets of influentials. This would reveal if certain influential group(s) on Twitter would also be influential in other ways. In particular, I wanted to identify influentials who were both actively disseminating information to the public (i.e., in TT or TR groups) and passively receiving attention from the general public on social media (ie, in TRRT or TM groups).

2.6 Results

2.6.1 Descriptive Results of the Zika-Related Tweeting Dynamic

A total of more than 6 million English tweets with the keyword Zika posted during 2016 were retrieved, of which approximately 4 million were original posts, and the remaining were RTs. More than 70% of the original posts received no retweet at all, and only 2% of tweets received at least five retweets. The Gini coefficients of the number of retweets were 0.74 and 0.98 for all original tweets and original tweets that received retweets, respectively. This indicated a very high heterogeneity in the potential influence of individual tweets on social media.

2.6.2 Association between Sporadic Critical Events and Zika Tweeting Dynamics

The peaks of Zika tweets were not synchronized with peaks of Zika counts, as discussed in the previous section. A large number of Zika-related tweets were associated with a few sporadic real-world events.

The association between Zika-related tweets and the unplanned real-world event WHO-PHEIC (Public Health Emergency of International Concern) announcement was shown in Fig. 2.2. WHO and WHO-PHEIC tweets were subsets of all Zika-related tweets. Upper panel of Fig. 2.2. was for the absolute number of tweet counts. The blue signal showed that all Zika-related tweets in 2016. The green and orange ones represented WHO and WHO-PHEIC signals, respectively. The lower panel of Fig. 2.2. showed the percentage of WHO and WHO-PHEIC tweets relative to all Zika tweets. If a tweet had both keywords/keyphrases of WHO and PHEIC, then the same tweet would be included in both categories. PHEIC and WHO related tweets had a high overlap (>50%), indicating the substantial impact of the WHO PHEIC announcement on public discourse on social media.

The keyword "WHO" had a strong presence in Zika tweeting throughout the first two quarters of 2016. There was a sudden rise in the number of tweets between day 31 and

32 of 2016 (Fig. 2.2); the number of Zika-related tweets increased drastically from 1481 on day 31 (Jan. 31) to 21171 on day 32 (Feb. 1), when WHO announced Zika epidemic as PHEIC. On Feb. 1, 2016, 35% of all Zika-related tweets were relevant to WHO, and 27% were about the announcement of PHEIC. This announcement also caused cascading public announcements in countries such as Brazil, Honduras, and the U.S. The highest number of tweets (92,000) posted on a single day regarding Zika was observed on day 34, just two days after the WHO-PHEIC announcement. Therefore, the unplanned WHO-PHEIC announcement was the driving force of the largest peak of Zika tweeting dynamics in 2016. It is worth noting that the discussion about the PHEIC started on Jan. 28, when the director-general of WHO announced that she convened the International Health Regulations (IHR) emergency committee and would have a meeting on Feb. 1 [86]. In addition to this peak, WHO-PHEIC signal had another prominent peak around day 323 (Nov. 18, 2016; Fig. 4). On Nov. 18, 2016, about 32% of the Zika tweets were related to WHO-PHEIC because WHO declared that the Zika epidemic was no longer a PHEIC on that specific day. Therefore, our EventPerisope analytical pipeline was effective in identifying and evaluate the impact of real-world events on tweeting dynamics.

The association between planned event RIO2016 and the peaks of Zika tweeting were shown in Fig. 2.3: The upper panel was for the absolute number of tweet counts. The blue and green signals showed all Zika-related tweets and RIO2016 Olympics tweets in 2016, respectively. The lower panel showed the percentage of RIO2016 tweets relative to all Zika tweets. In general, discussions about Zika and RIO2016 Olympics started from the beginning of 2016 all the way through a few days after the Olympics ended. In other words, although the event of the RIO2016 Olympics only lasted for two weeks, discussion of this event with regard to Zika went on throughout the entire year because the Olympics was a planned event. Specifically, on its opening ceremony day (Aug. 5), 12% of all Zika tweets were related to RIO2016 and up to 18%

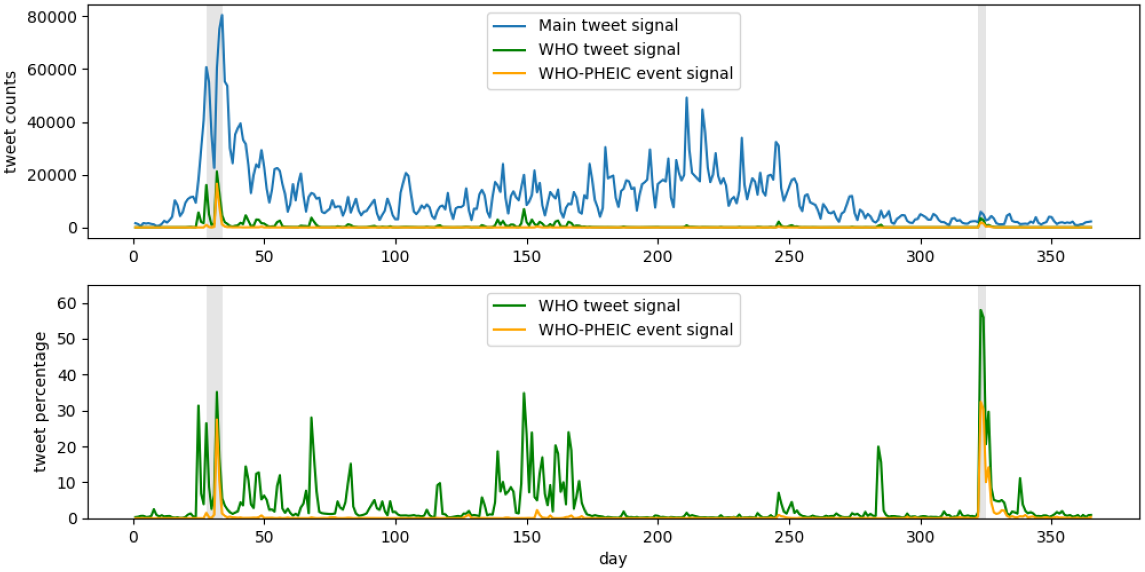


Figure 2.2: **Upper Panel:** The blue curve depicts Zika-related tweet counts in 2016. The green and orange curves represent the number of tweets containing WHO and WHO-PHEIC keyphrases, respectively. **Lower Panel:** Green and Orange curves show the percentage of tweets containing WHO and WHO-PHEIC keyphrases, respectively.

in the next day. In addition, RIO2016 had a prominent presence in other noticeable peaks of the Zika tweeting signal. For example, RIO2016 constituted 71% of all Zika related tweets on day 149 (May 28). Our further investigation revealed that on day 133 (May 12), a researcher started the debate that RIO2016 should be canceled or

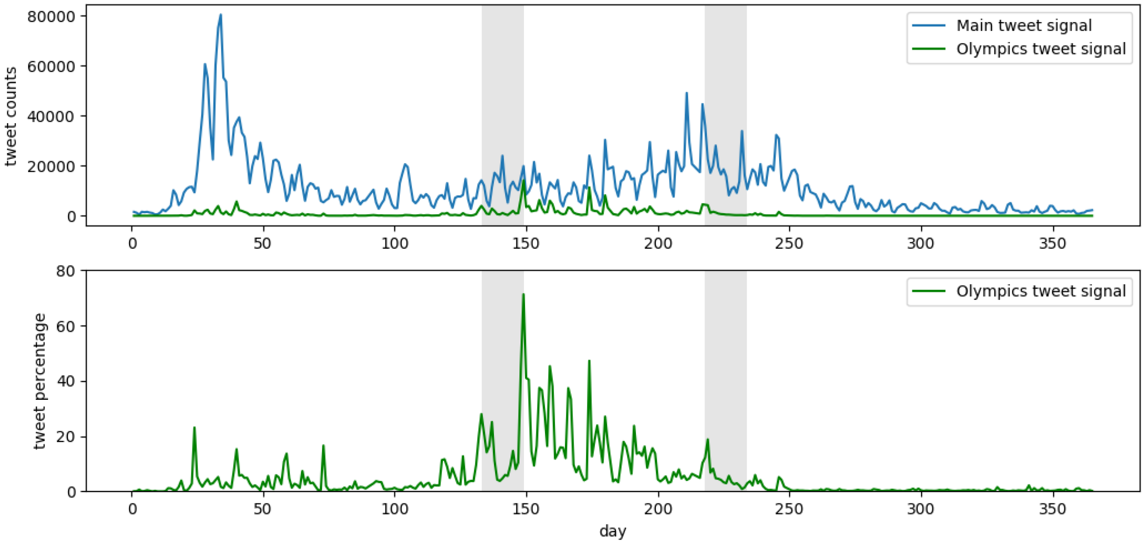


Figure 2.3: The green signal in the lower panel represents the percentage of tweets related to Rio 2016 Olympic Games.

at least postponed amid concerns of the Zika outbreak [87]. However, on day 149 (May 28), WHO released a statement [88] explaining it was not necessary to take such an action. Because of the WHO announcement regarding RIO2016 and Zika on day 149, the WHO-Zika signal also had a peak on day 149; WHO-related Zika tweets comprised 34% of all Zika tweets (Fig. 2.2). These results supported hypothesis H1 that Zika tweeting dynamics were triggered by other events in the real world.

2.6.3 Association between Online Influentials and Zika Tweeting Dynamics

In this section, I presented the role of TT (top tweeter), TR (top retweeter), TRRT (top received retweets), and TM (top mentioned) influential user groups, as defined previously

2.6.3.1 Comparison between Each Group of Influentials and Zika Tweeting Dynamics

Tweeting dynamics in TRRT, TT, TM groups, and retweeting dynamics in the TR group were extracted and constructed for the top 100 users in each group. Quarterly association between these groups' tweeting dynamics and overall Zika tweeting dynamics were shown in Fig. 2.4-2.7. Each figure had three panels. The upper panel showed the overall tweeting dynamics, the middle demonstrated the tweeting dynamics of the particular influential group, and the bottom one showed CCF of the two signals. Group-level tweeting dynamics in TT, TRRT, TM were highly correlated with and approximated the shape of the overall tweeting dynamics (Fig. 2.4, 2.6, 2.7). However, the retweeting dynamics of the TR group was not closely associated with overall Zika tweeting dynamics (Fig. 2.5). In the TR group, there were peaks in their retweeting signal on day 170, 173, 265, and 303; however, no noticeable corresponding peaks were identified around these days in the main Zika tweeting signal. I conjectured that TR group, in general, would be more active following certain undetected events, which did not necessarily coincide with the overall Zika tweeting dynamics.

More importantly, for TT, TRRT, and TM groups, the maximum CCF occurred at +1 day lag in the first three quarters of 2016 (Fig. 2.4, 2.6, 2.7), indicating that these groups' tweeting activities were one day ahead of the overall discussion on Twitter. For example, the peaks in the overall Zika tweeting signal were lagging behind the peaks in the TM group by approximately one to two days. Therefore, these influential groups' tweeting activities were not only highly associated with the overall tweeting dynamics, but these influentials were also the potential driving forces of the overall Zika discussions on Twitter. As a result, by observing a few hundred influentials' tweeting activities, I could accurately predict the upcoming rise and fall in the overall tweeting dynamics. Nevertheless, such lag diminished to zero in the fourth quarter for all three influential groups, as the Zika epidemic and PHEIC ended in the fourth quarter of 2016.

In addition, I examined the contributions of individual users in each of these influential groups TT, TRRT, and TM. I further calculated the CCF between a user's tweeting time series and the overall tweeting dynamics in each quarter as well as the entire 2016 (Fig. 2.8). Time lags of the majority of influential users were very close to zero, which implied that these users could not be driving the overall discussion of Zika on Twitter, but rather participating in the discussion. However, there were a few users whose time lags were substantially positive, indicating their potential role in driving the overall Zika tweeting dynamics. Furthermore, the quarterly results revealed tweeting dynamics of influentials at finer temporal resolution than yearly results (Fig. 2.8). Note that in each panel, the first four boxplots (labels 1-4 in x-axis) were quarterly, and the last one (5) was for the entire year of 2016. In general, most influentials did not engage in Zika discussions on Twitter constantly and continuously across the entire year of 2016. They might be active and highly influential during certain time periods when they were interested in Zika hence participated in discussions on Twitter. As a result, aggregating all individual influential users' tweeting activities in the entire year would undermine each user's temporal dynamics of tweeting and consequently, its time-specific influence on the overall discussion dynamics on social media.

2.6.3.2 Overlap between Influential Groups

In addition to exploring each potential influential group's role in driving the Zika tweeting dynamics, I also investigated whether different influential groups had overlaps. Table 2.1 showed the year-long intersections between any two groups of influentials, while Table 2.2 was on a quarterly basis for selected groups. TM group had no member who also belonged to TM or TRRT groups, and TT group had no intersection with TRRT group. These results suggested that being highly active did not necessarily guarantee to receive a lot of mentions and/or retweets from other users on social media. Therefore, active and passive influentials about the emerging health issue of Zika on social media were distinctive users. On a quarterly basis, there were quite a few influentials being mentioned and being retweeted a lot at the same time (Table 2.2 column 2). On the other hand, there were only a few users in TR group who were also highly mentioned and retweeted (Table 2.2 column

Table 2.1: Number of intersections between the four groups of users

	TM	TRRT	TT	TR
TM		47	11	0
TRRT	47		0	0
TT	11	0		6
TR	0	0	6	

Table 2.2: Quarter based intersections between the four groups of users

Quarter	TM-TR	TM-TRRT	TR-TRRT
1	4	49	3
2	4	43	5
3	3	44	2
4	6	45	8

1 and 3). These user accounts belonged to health organizations, such as @cdchep, and @CDCChronic, and also a few well-known but independent individuals, such as @Laurie_Garrett and @MackayIM. This reinforced our previous finding that active and passive influentials were not the same users. For public health agencies such as CDC, while they might actively disseminate information to the public on social media, their efforts were not well recognized by the general public users. Therefore, health agencies needed to craft more effective strategies to engage the public with the discussions of an emerging health issue on social media.

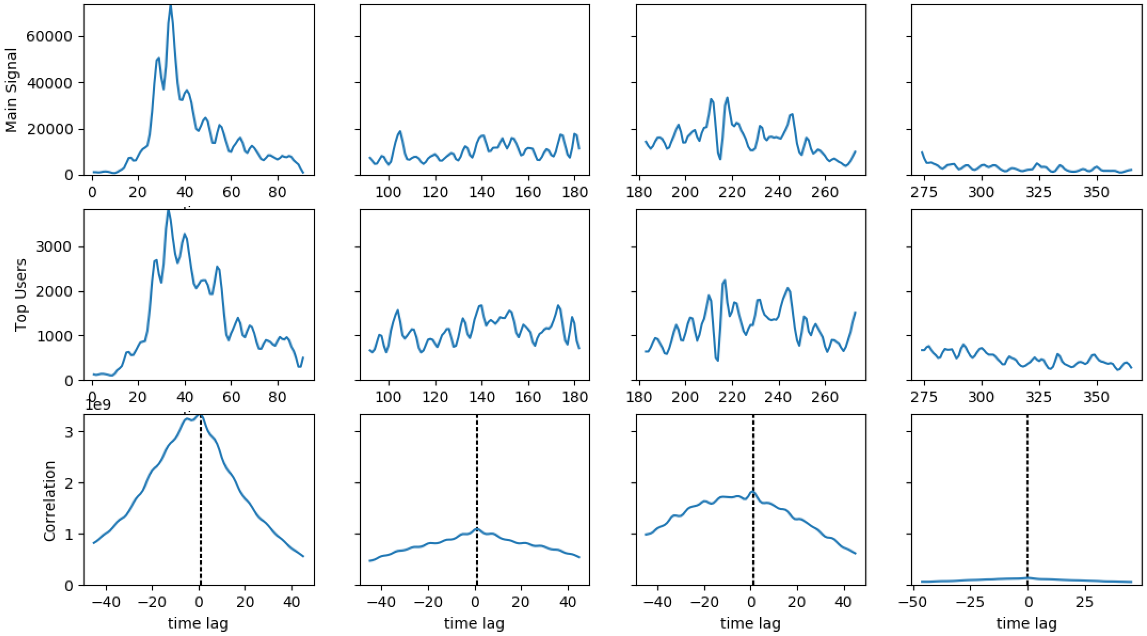


Figure 2.4: Correlation between tweet stream of all users and users in TT group.

2.7 Discussions

Communicating with the general public is essential in risk communication during public health emergencies [89]. Hosting a large and diverse population, social media platforms such as Twitter are valuable resources for public health professionals to understand and analyze public opinions on emerging health issues [80, 90, 91, 92, 93]. During the 2016 Zika epidemic, Twitter was demonstrated to be an ideal place to explore public concerns and interests about the disease through time and across different

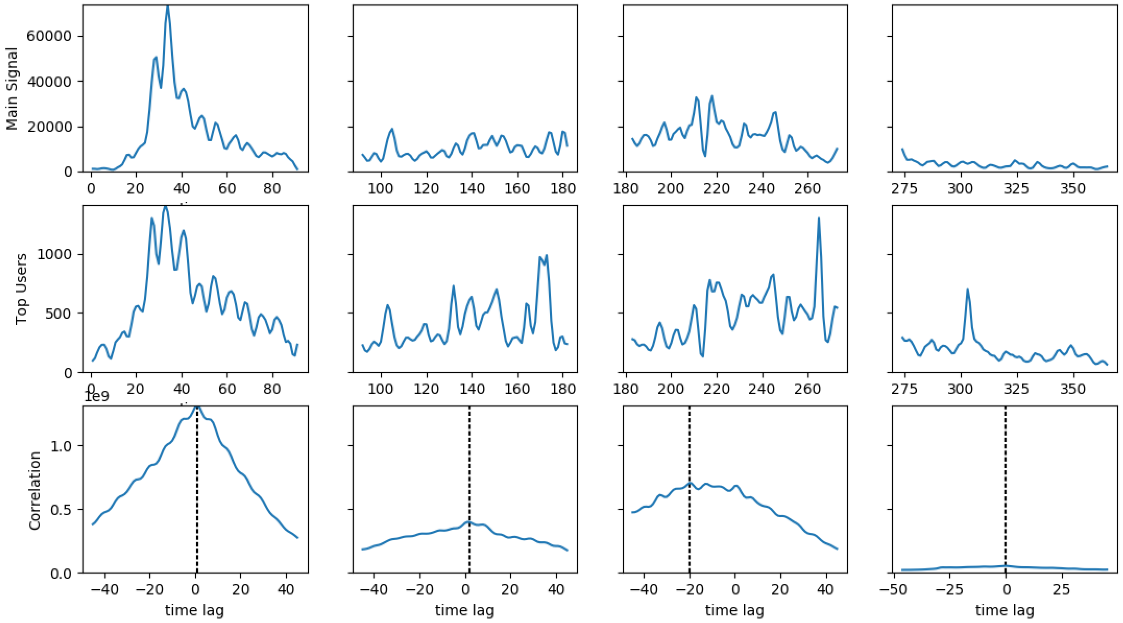


Figure 2.5: Correlation between tweet stream of all users and share activity of users in TR group

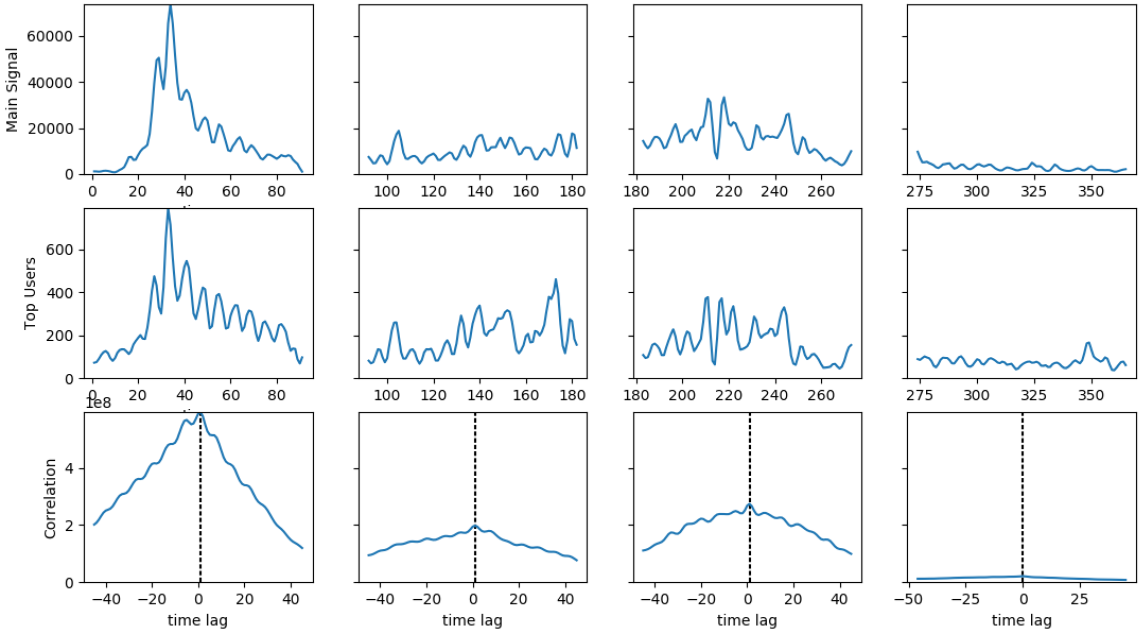


Figure 2.6: Correlation between tweet stream of all users and users in TRRT group.

locations [91, 76, 94, 69, 95, 78]. In addition to a means of understanding public opinions, social media platforms are utilized by health professionals to communicate with the public and to disseminate accurate and timely information regarding an ongoing

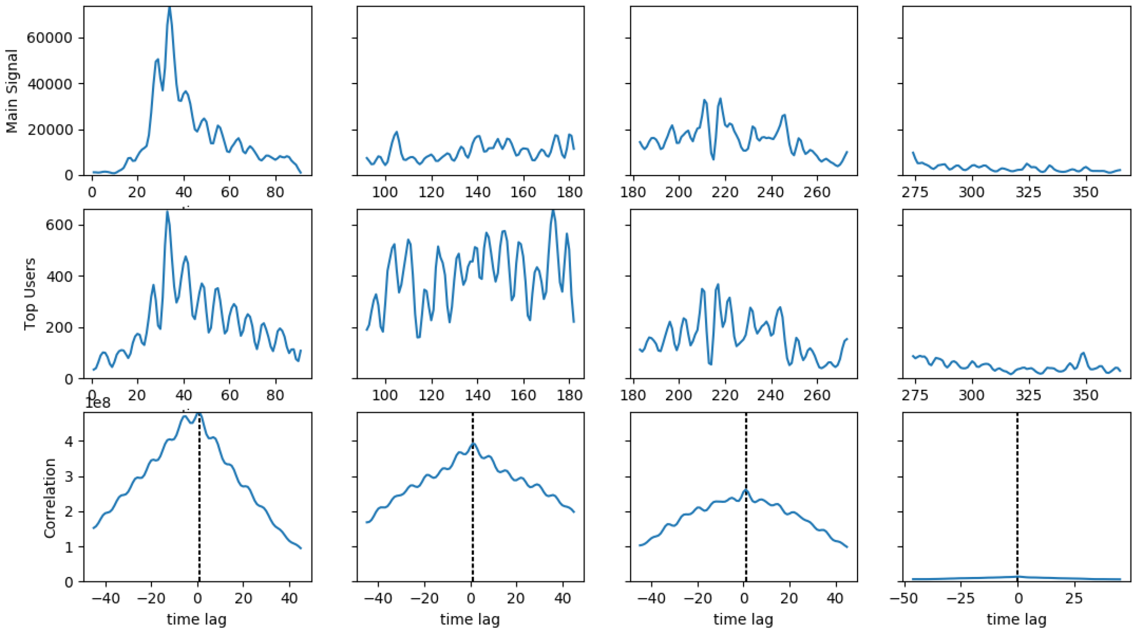
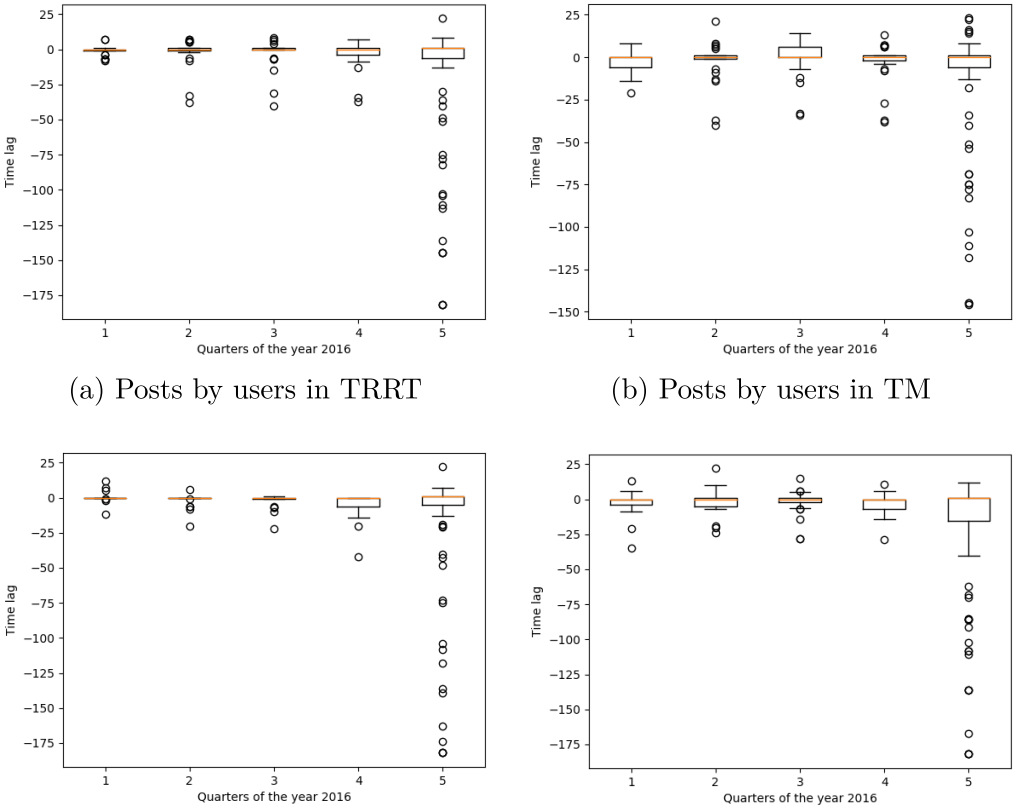


Figure 2.7: Correlation between tweet stream of all users and post activity of users in TM group.

health emergency [96]. For example, in [97] Chen et al. has evaluated the role of CDC in disseminating Zika-related information on Twitter during the Zika outbreak. That study revealed that CDC played a critical role in tweeting Zika-related information during the first quarter of 2016 when the actual disease counts were still relatively low. However, CDC's Zika-related tweets quickly and drastically decreased after the first quarter of 2016, when the Zika case counts went up [69]. One important yet under-explored aspect of online discussions of health emergencies is to identify potential driving forces that can lead and change the dynamics of discussions on social media. Identifying such influential factors/contributors is critical for devising effective strategies in health crisis management and risk communication. Currently, studies have shown that discussion on social media can help estimate disease burden more accurately, known as the Infodemiology [79, 98, 99]. However, it is not clear whether and how the actual situation of a health issue influences the public's perception and discourse on social media. Moreover, the correlation between real-world events and their potential impact on online discussions of health emergencies is not well investi-



(a) Posts by users in TRRT

(b) Posts by users in TM

(c) Posts by users in TT

(d) Posts by users in TR

Figure 2.8: Boxplot formed from comparing the tweet activities of individuals in four groups with the main stream

gated and understood. In addition to investigating the impact of "what happened," it is critical to evaluate the role of online influential actors, i.e., who would be the online opinion leaders that drive online discussions. These two research questions correspond to the two hypotheses that I have investigated in this study. My systematic and comprehensive analyses have provided a novel and holistic view of different factors impacting discussions about a health emergency on social media. This new perspective will help us better understand the complexity of such discussions.

In the future, there are a number of directions that I could pursue to further improve and expand this work. As one example, the last hypothesis that investigates the role of online influentials is not mutually exclusive from the first hypotheses. For instance, my preliminary study has shown that during and immediately after the WHO-PHEIC announcement on Feb. 1, 2016, many news agencies' Twitter accounts were helping to disseminate this announcement on Twitter. Therefore, both the critical real-world event (WHO-PHEIC announcement) and online influentials (news agencies' accounts) were simultaneously driving Zika tweeting dynamics. In the future, I plan to further explore changes in the dynamics of discussions by constituent contributors. In this chapter, I have demonstrated high association and temporal precedence between the tweeting activity of influentials and the overall tweeting dynamics. Online influentials' tweeting signals preceded the overall tweeting signal regarding Zika, which were strong indicators of potential causality. My results suggest that tweeting activities of TRRT (top received retweets), TT (top twitterer), and TM (top mentioned) groups, are good representatives of the overall tweeting dynamics. Therefore, their tweeting dynamics can be used to accurately approximate overall discussion dynamics on social media and to further predict the upcoming changes in discussion dynamics effectively.

In order to investigate discussion dynamics on Twitter, a highly sophisticated and complicated social media platform with millions of tweets, I have utilized an array of different computational methods, including time series analysis, signal processing, content analysis, and information theory computations. In particular, I have developed an analytical pipeline, EventPeriscope, to integrate and consolidate these different computations. EventPeriscope pipeline is the practical outcome and contribution of this study. Comparing to other similar analytical frameworks, EventPeriscope has the advantage of detecting both planned and unplanned events related to a specific discussion topic. This analytical pipeline can be readily transferred and applied to investigate other emerging or non-emerging issues on social media such as general discussion of health issues, identify potential driving forces of the discussion, and evaluate their influence.

I need to point out that the two major drivers on tweeting dynamics in this study are not an exhaustive list of possible drivers. Further potential drivers, such as individual user or organization user, verified or unverified user status, will also be investigated in the future. In addition, EventPeriscope can be used to detect other concurrent issues that might also influence Zika tweeting dynamics, such as the 2016 U.S. Presidential Election.

2.8 Conclusion

This chapter analyzed Zika-related tweeting dynamics in 2016 when Zika became a global concern. I revealed potential drivers of Zika discussions on social media by testing three hypotheses. First, I showed that peaks of Zika tweeting dynamics were significantly influenced by and associated with critical real-world events, both planned, such as the Rio Olympics and unplanned such as the WHO-PHEIC announcement. I further evaluated the role of potential online influentials and demonstrated that top twitterer (TT), top mentioned (TM), and top users whose tweets were retweeted many times (TRRT) were potential drivers of the overall discussion of Zika on Twitter. Through these careful analyses of tweeting dynamics, my study revealed potential contributors and drivers of discussion on an emerging health topic. Insights gained from this study could be applied to other emerging health topics in the future. More importantly, I demonstrated the feasibility of my comprehensive analytical approach and the EventPeriscope framework to investigate online discussions dynamics of health emergencies and to identify potential driving forces of these discussions.

CHAPTER 3: Comparing Health-Related Misinformation and Relevant Information Dissemination on Social Media

3.1 Overview

Health-related misinformation infiltrates and proliferates on social media. They cause unnecessary confusion, anxiety, mistrust, and anger in society, especially during health emergencies. Current studies generally focus on information content aspect to detect potential misinformation, while ignoring other important aspects of information. In this chapter, I aim to provide a novel methodological framework to distinguish health misinformation from relevant information on social media based on information dissemination dynamics. Information dissemination of a specific tweet is defined as its retweet (RT) process, i.e., who-retweeted-whom. I aim to extract new features from information dissemination dynamics that will shed light on understanding why and how health misinformation proliferates, facilitating the development of a more accurate health misinformation detector, and guiding more effective health communication strategies in social media era.

3.2 Introduction

A growing number of users from a wide demographic rely on social media platforms as a real-time source of information to share and obtain news about various topics [100, 101]. These platforms are fast, omnipresent, and easily accessible. As a result, a piece of information can go viral in a short time. Contents that are posted on a social platform, such as Twitter, are mainly user-generated and the lack of effective systematic fact-checking mechanisms makes these high-velocity environments susceptible to be abused for the propagation of misinformation. In particular, the proliferation of health-related misinformation on social media, especially during health emergencies, is a serious threat to modern societies [102]. Anti-vaccine debates on Twitter and Facebook [5, 103, 104], misleading videos about Zika on Facebook during the 2016 Zika pandemic [6], or misleading videos about tobacco, vaporing, and marijuana products on YouTube [105] are a few examples showing the necessity of developing a comprehensive surveillance system to identify and neutralize the effect of different forms of health-related misinformation. Social media platforms have been proven to be an effective tool to enhance the public's knowledge of health [106, 107] and also a rich resource to study the public's perspectives, reactions, and concerns toward various topics [108, 109, 68, 110]. Equipped with an effective surveillance system, practitioners can closely monitor discussions to understand the uncertainty and lack of knowledge in a particular area and therefore act proactively by providing more resources[111, 112, 113, 114]. On the other hand, they can detect misinformation and neutralize it by providing true information [7]. Developing such a system is not a straightforward task and demands the collaboration of researchers from different disciplines [115].

Most research works on misinformation generally focus on the content aspects, linguistic features, and motivation of its propagators [38, 116, 117, 118, 119, 120, 121] to detect and characterize potential misinformation while paying less attention to other important aspects. Since a misinformation epidemic very much resembles real infectious disease epidemic, focusing the misinformation content is similar to working on the pathogen side alone. Misinformation content can be altered to resemble real information to avoid being detected by automated algorithms [50], similar to a pathogen that can mutate to avoid being detected by the immune system. Therefore, I suggest that only relying on textual content is not adequate to comprehensively tackle the health misinformation challenge, and we need to understand health misinformation from more aspects.

The pathogen, the environment, and the hosts are collectively known as the inseparable *Epidemiological Triad* or *Epidemiological Triangle*, a fundamental concept in infectious disease epidemiology. Similarly, I propose that the misinformation, the online, especially social media environment, and the users also form an *Infodemiological Triad*. One of the key approaches to investigate an infectious disease outbreak is to track the trajectory of epidemic, i.e., who are infected from whom at what time. In this study, I define the dissemination of a particular piece of (mis)information as a dynamic process where the original post (e.g., a tweet) is propagated on an information dissemination network [122, 123, 124]. There are three common ways that a piece of information (post) can be disseminated on social media: through retweet (reposting or sharing in other social media platforms), liking, and commenting. It is technically difficult to track likings from users as most social media platform's metadata do not contain explicit information of who liked whom at what time. Commenting, on the other side, adds commenters' own opinions on top of the original post, which may or may not necessarily align with the original one. Retweeting shows that the retweeter, if not a bot or cyborg, cognitively and actively recognizes the importance of the original post and is willing to let other users see their retweeting activity because the record of retweet will show up on the retweeter account, but liking a post will not. Therefore, in this study, I focus on retweet as the major information dissemination method, where the temporal dynamics of retweets occur on the network of retweeters [124, 50]. I aim to develop an algorithm to infer and reconstruct the information dissemination network through retweeting activities, extract quantitative features of the networks of both misinformation and real information groups, and investigate how they differ quantitatively using network analysis techniques.

To perform my analysis, I especially focus on the Zika epidemic crisis in 2016. During the Zika outbreak in 2016, Centers for Disease Control and Prevention (CDC) had a strong presence Twitter by releasing the latest findings and instructions [94, 69].

However, uncertainty about the root cause and transmission rout of this virus opened the room for the proliferation of rumors and misinformation [71, 72].

The outcome of this study is a novel set of features extracted by comparing and contrasting misinformation and real information cascades. This study will leverage our understanding of how specific health misinformation proliferates on social media and how they might have outcompeted real information. Eventually, the insights from this project will help develop more effective health communication strategies in social media against various health-related misinformation.

3.3 Method

3.3.1 Data Retrieval and Processing

The entire year of 2016 (January 1 - December 31, 2016) was chosen as the sampling period for this study. This decision was made for the following reasons. First, the Zika epidemic did not become a nationally notifiable condition in the United States until 2016, although some suspected Zika cases were reported in the United States at the end of 2015. Second, this time period covers the major milestones in the Zika epidemic timeline, such as the WHO's initial warning predicting the spread of the Zika virus across the Americas, the official declaration of the PHEIC for the Zika virus on February 1, 2016, and the end of the PHEIC on November 18, 2016. Using Zika as the keyword, a total of 3.7 million English tweets and retweets published in 2016 were collected via the Gnip application programming interface (API) through the university's data science program. This dataset was the complete dataset, including all English Zika discussions occurring on Twitter in 2016. It was not the common 1% sampled data from Twitter's own API. Therefore, this dataset provided a more comprehensive, complete, and less biased view of the public discourse of Zika on Twitter in 2016.

3.3.2 Misinformation Identification and Relevant Information Matching

All the original Zika tweets were ranked based on the number of received retweets, from highest to lowest. Following this descending rank, the top 5,000 most retweeted Zika tweets were selected as the sample pool. Then an operational definition of misinformation was established such that the information in the tweet was not evidence based. Peer reviewed journal articles and conference proceedings, government and health agencies (e.g., CDC and WHO) announcements and statistics, fact-checking websites, were all used to evaluate the validity of the tweet. Based on this definition, misinformation included but was not limited to the following categories: misconception or misunderstanding of scientific concepts were those that could be verified by current literature, for example, stating vaccination as a cure instead of prevention of Zika. Fabricated stories had no scientific or actual real-world evidence, which could also be verified by peer-reviewed literature and health agencies' statistics. Conspiracy theories related Zika with unverifiable yet controversial topics such as genetically-modified (GM) mosquitoes, "big-pharma", and vaccinations. Similarly, rumors were information which could not be verified or lack of evidence support. Discrimination and bias were using Zika to discriminate against a group of people of certain gender, race, ethnicity, nationality, religious belief, political view, and sexual orientation. Hate language, name-calling, and profanity, on the other hand, was usually towards a specific person but not the larger group. Disinformation did not explicitly discuss Zika but rather used Zika as a hot and controversial topic to divert the public discourse to other topics that the propagators were interested in, for instance, GMOs. Sarcasm, while generally not considered as a type of misinformation, had the ability to hide the real intention and confuse the readers, especially in this fast-paced social media age. Therefore, specific sarcasm tweets could also spread the potential misinformation within it. Note that not all sarcastic tweets were misinformation, and sarcasm tweets were treated case-by-case. An example of these different types of misinformation is shown in Table 3.1.

Table 3.1: Examples of Different Types of Zika Misinformation Tweets

Types of Misinformation	**Content**
Misunderstanding/ Misconception	University of Pittsburgh finds vaccine that cures Zika in mice. What's up every other University? Does Pitt have to cure every disease?
Fabricated Story	Zika virus: Health department urges calm, says no cases... https://t.co/e6SaiYp9UP #Zikavirus
Conspiracy	"Zika Hoax Created By The World Health Organisation To Destroy Brazil's Economy!" https://t.co/WZqj1nQagV https://t.co/NEDTLv7oqE
Rumor	Is the dreaded Zika virus another giant scam? ...Explain the small-head babies then! https://t.co/bKND4aZWRz https://t.co/jAaYmavx55
Discrimination and bias	@col_nj #Illegal; #Refugee INVADERS carrying BIOLOGICAL WEAPONS IN bodies too. #tcot#TB#Zika#Ebola#MERSA#Chagas#Dengue #Leshmaniasis
Profanity	Draymond Green punched Harambes **** and gave him Zika. Please RT
Disinformation	Zika Outbreak Epicenter in Same Area Where GM Mosquitoes Were Released in 2015: https://t.co/IHJ3M1wof8
Sarcasm*	Scientists Confirm First Case Of Zika Transmission From Article To Reader https://t.co/7peBpSSVlw https://t.co/M6paYI08jd

Note that these categories were not mutually exclusive. For example, a Zika tweet could be both disinformation and conspiracy theory at the same time. In addition, it was not uncommon for a single tweet to have multiple components of facts, statements, and opinions. In this circumstance, if any component was identified as not evidence-based, then the entire tweet would be classified as misinformation, even if other components may contain real information. Mathematically, if a tweet X could be divided into n components, then the validity of entire X is determined by the logic AND operator on all its components such that $X_1 \wedge X_2 \wedge ... \wedge X_n$.

I acknowledged the fact that there was currently no consistent and universally accepted definition of misinformation across disciplines, and sometimes people used similar terms such as misinformation, disinformation, rumor, "fake news" interchangeably without a clear distinction. As such, in this study, a clear and operational definition of misinformation based on the evidence in the content was provided. This definition of (health/medical) misinformation was in accordance with the definition of Evidence-based Medicine (EBM) that health professionals were familiar with.

After the definition of misinformation about Zika was established, two researchers on the team who had experience with this topic independently checked the validity of the same set of 100 randomly selected tweets from the 5,000 sample pool. The initial intercoder reliability κ was 0.86. Several rounds of discussion and revision were made before κ reached 1.00 (no disagreement) as the desired threshold value. In addition, multiple experienced researchers in the fields of arboviruses, vector-borne diseases, and infectious disease epidemiology were consulted about their opinions on the contents to ensure the validity and operationalization of the definition of misinformation. Afterward, the validity of all remaining most retweeted Zika-related tweets was evaluated and the top retweeted tweets with misinformation in them were identified. The set of all misinformation tweets in this study formed the most retweeted misinformation group.

Because the definition of misinformation was exclusive, i.e., any tweet that was considered non-misinformation was containing real information. In order to more accurately evaluate the differences between real and misinformation about Zika on social media, a comparable group of tweets with real information were identified. For example, the severity of the Zika epidemic in 2016, which could be summarized as the number of new cases in a given day, could heavily influence the public engagement of Zika discussions on social media. The matching method with the following two criteria was used: 1) the tweets with real information published within the plus/minus 3-day window of the identified tweets with misinformation, 2) the tweets containing real information with similar numbers of retweets (plus/minus 10 retweets) compared to those of the tweets with misinformation. This matching method was intended to minimize influences from potential confounders. By controlling these confounding factors, one should be able to get a more accurate answer on whether information validity actually influenced retweeting activity on Twitter. The identified tweets formed the real information group.

For each identified misinformation and real information tweet, its metadata, as well as metadata of all its retweets, including posting date/time, tweeter/retweeter IDs, friends/followers information, etc. were also collected. This information was critical to track information dissemination through retweeting, construct information dissemination networks, and conduct statistical analyses to quantify differences between the misinformation and real information groups.

3.4 Constructing Dynamic Information Dissemination Networks of Retweeting

Information dissemination is defined as a process in which a piece of information (e.g., a tweet) is propagated among a network of users (retweeters) within a specific time period [122]. Therefore, two main attributes associated with information dissemination are time and users who propagated them. In the context of Twitter, a piece of information is incorporated in a tweet and retweeters of that tweet are the spreaders. Therefore, to characterize information dissemination corresponding to a tweet, I consider two main attributes: 1) the time points when retweeters relay the tweet by retweeting it, 2) follower-followee relationship among the retweeters of the tweet. To investigate the flow of information among the users, I estimate a diffusion network based on the follower-followee relationship between the retweeters and the timestamp of their retweets.

Mathematically speaking, a network (or graph, G) has two sets of elements: a set of nodes (or vertices, denoted as V) and a set of edges, denoted by E, where an edge connects a pair of vertices v_i and v_j. If the pairs of vertices connected by edges are ordered, then the graph G is called a directed graph. As an example, in a retweeting network, the node v_i follows the node v_j but v_j does not follow v_i, then there is a directed edge from v_j to v_i indicating the information dissemination direction through retweeting. A graph can be edge-weighted; that is, every edge e in E is assigned by the weight $W(e) \in R$. A weighted graph is represented by $G = (V, E, W)$, where $W : E \rightarrow R$. Therefore, I use a directed and weighted graph to model the follower-followee relation between retweeters who take part in the propagation of a tweet. Directions in such a network show the friends-followers relation between the propagators of the tweet.

On top of this user network, I then construct an *information dissemination network* to infer how the information flow among these users (retweeters) over time. First, to show the flow of information among nodes, I reverse the direction of edges in the followers-followee network. As a result, the in-coming edges to a node show potential sources that it could receive the information (tweet) from. Next, for every edge e_i, I set the weight to be the time difference between the retweet time of vertices which are connected by e_i. I will then continue by making two assumptions. First, an actor who retweets a post has seen it because at least one of her neighbors that she follows has retweeted the post before her. Thus, I discard edges between her and her other neighbors (followees). Isolated vertices within the network are also dropped because, obviously, they do not have any contribution to information dissemination. Second, in the flow of information, I assume that a node is influenced by one and only one of her followees. Therefore, the resulted information cascading network is a tree. Directions in such a tree is from followees to their followers to show that for a retweeter, who could be her potential source of information. When a Twitter user opens her page, she sees posts by her followees ordered chronologically. If two of her followees have retweeted the same tweet, she sees the most recent one first. Therefore, to determine the path that a tweet takes to reach a retweeter, I assume that her followee who has retweeted the tweet most recently is the one that has influenced and triggered her to retweet it. To construct the information dissemination tree based on this logic, I use the minimum spanning (MST) algorithm. In practice, time is not the only factor that influence a user and make her to retweet a post. However, for the purpose of this study, which is showing the propagation pattern of misinformation contained tweets is different from those carrying real information, this assumption suffices. Moreover, MST let us capture the minimum time that a tweet took to get viral.

3.4.1 Computing and Interpreting Important Network Metrics Relevant to Information Dissemination

Once the networks are constructed for each tweet, important network metrics that are highly relevant to information dissemination are computed and compared both within and between the misinformation and relevant information groups.

In this study, I extract a total of nine network metrics, including network reach, network influence, network diameter, network density, network modularity, Wiener index, structural virality, top out-degree centrality score, and top betweenness centrality score. These metrics quantified and characterized network structures from different aspects and across scales in the network. Below, I provide a succinct description of these metrics

Network Reach (REA) of a dissemination network measures the number of unique vertices, i.e., *unique* tweeter/retweeter IDs in this network, whereas *network influence* (NIF) represented the number of retweets. Both REA and NIF are considered as global (overall) properties of a network and quantify user engagement of the tweet through retweeting activity. Larger network reach and influence indicates more user engagement of the tweet. Note that the reach and influence of a given network are not necessarily the same, as a user could retweet the same tweet multiple times. Mathematically speaking, NIF is the upper bound or limit for REA. If each retweeter (user) retweeted exactly once, then REA is equal to NIF. If there is a large discrepancy between NIF and REA for the same network, it indicates that a few retweeters are trying to retweet and propagate the same piece of information multiple times. This behavior could be identified as a potential indicator of intentional amplification of the piece of information carried by a tweet.

Network diameter (DIA) is calculated as the shortest distance between the two most distant vertices (users) in the network. DIA represents the linear size of a network and therefore is also considered as a global (overall) property of a network. In general, the smaller the network diameter, the fewer steps information was required to pass through to reach distant users. Oftentimes, fewer steps indicate faster information dissemination from the original tweeting node, assuming information dissemination speed was relatively constant. Therefore, for effective and efficient information dissemination, a retweeting network with a smaller diameter would be more desirable. In addition, based on the minimum spanning tree method that I use to construct the dissemination networks, a small DIA shows that retweeters are more closely related to the source of information, i.e., the original posting user, and consequently are more influenced by the original poster of the tweet rather than other retweeters.

Structural virality (VIR), similar to DIA that quantified distance among vertices, measures the average distance between all pairs of vertices in the dissemination network [125]. The larger value of VIR indicates that the retweeters are, on average, further apart in the information dissemination network. An effective information dissemination network generally has a small VIR value.

Both DIA and VIR quantify distance among vertices in the network, the difference is that DIA focused on the shortest path while VIR is for the average distance. A large VIR can also indicate that within an information diffusion network, actors do not highly share their source of information (do not have the same source of information) *Network density* (DEN) measures the proportion of potential relationships that actually exist in the network. DEN is calculated by the total number of existing edges over all potential edges in a network, where an (a, b) edge with an arrow from a to b shows that the node b follows the node a and also a has retweeted the same tweet before b; thus, a could be a potential source of information who influenced b to retweet the tweet. Density is considered as another global-level network property. This quantity always takes values between zero and one. Density is equal to zero for a trivial graph in which there is no edge between vertices. In a clique or completely connected network where each node is connected to all other nodes within the network has network density equals one. Network density focuses on the relation of nodes within the network and aims to capture how they are intermingled. Thus, to calculate the network density, rather than the information dissemination network, I consider the network of retweeters which is constructed based on their followee-follower relationships and also the time difference between their retweet. The information diffusion network which is constructed on top of that followee-follower network is a tree and the number of edges within a tree always equals to $n-1$, where n is the number of nodes within the network. As a result, calculating network density for the information dissemination network cannot give us a deep sense of how nodes are interconnected.

Considering the assumption based on which I construct the follower-followee network, each node could only be connected to nodes who have retweeted a tweet before them. The potential number of edges that could be present in such a network with size n is equal to $\frac{n(n-1)}{2}$. Therefore, I calculate the density by

$$N(E)/(n(n-1)/2),$$

where $N(E)$ is the number of edges within the network.

A large value of DEN generally implies that vertices are connected more frequently; therefore, the information should be disseminated more effectively.

Network modularity (MOD) measures the likelihood of dividing a complete network into potential clusters, i.e., subgroups within which nodes are highly connected, but loosely connected among subgroups. Modularity is a local property of a network. The larger the modularity value, the more likely the network could be divided into subgroups. For information dissemination through retweeting, such a subgroup could be explained as an influential drawing a lot of attention by attracting many followers to retweet. However, more subgroups might reduce the efficiency of information dissemination, as subgroups generally do not have a lot of interactions among themselves (e.g., many edges between subgroups), otherwise these subgroups would eventually consolidate into a larger (sub)group. Note that MOD is the only network metric in this study that could possibly have a negative value, and negative MOD indicated less local subgroup structure [126].

Wiener index (WIE) is defined as the sum of the shortest paths between all pairs of vertices. From an information dissemination perspective, a star-like network where all retweeter vertices retweet directly from the original posting vertex would have smaller WIE comparing to a more chain-like network, even if both networks had exactly the same number of vertices (NIF). Therefore, the smaller Wiener index value indicates that the network could have a more star-like structure.

Out-degree centrality (OUT) for a node v is defined as the number of nodes that have an incoming edge from v. OUT measures how much influence each retweet vertex has in terms of spreading the information further out. Intuitively, OUT is proportional to the number of outgoing retweeting edges of a given vertex, and larger out-degree centrality value indicates that the vertex has a lot of offspring vertices to further disseminate the information. I calculate the entire OUT distribution and present the largest OUT value of all retweeters in a network.

Betweenness centrality (BET) of a node is defined as the number of shortest paths that pass through the node. BET does not directly measure the influence of directly disseminating the information out but quantifies the importance of the vertex in terms of the connectivity of the network. A large betweenness centrality shows the vertex is critical for the flow of information in the network, and removing such a vertex could reduce or even completely block information dissemination. OUT and BET metrics focus at the vertex level, and quantifies the role of an individual vertex for information dissemination. Similar to OUT, I compute BET for all retweet vertices in a specific tweet diffusion network, and consider the largest one.

Note that these two types of centrality are generally independent of each other, large out-degree centrality does not necessarily imply large betweenness centrality score, and *vice versa*. In this study, for every tweet dissemination network, I demonstrate the largest BET value in all retweet vertices, and compare the resulted values between the misinformation and real information groups.

In summary, these network metrics comprehensively represent different aspects of networks at different scales, from the overall global network level to local cluster level and all the way down to individual vertex level. Although there are other network metrics, these nine measures can adequately operationalize the structures of the diffusion networks constructed in this study. Therefore, they can give a good understanding of the differences between the patterns of misinformation and real information propagation.

A descriptive summary of parameters corresponding to the diffusion network metrics in Table 3.2

3.4.2 Comparative Analysis of Misinformation and True Information Propagation

To identify different information dissemination patterns between misinformation and true information, I perform the Kolmogorov-Smirnov (K-S) test for each metric

Table 3.2: Network Metrics

Network metrics	**Description**
Network reach	The number of unique vertices i.e., unique IDs
Network influence	The number of retweets
Network diameter	The shortest distance between the two most distant vertices
Structural virality	The average distance between all pairs of vertices
Network density	The existing proportion of potential relationships
Wiener index	The sum of the shortest paths between all pairs of vertices.
Network modularity	The likelihood of dividing a network into potential clusters
Largest Out-degree	The largest Out-degree of all retweeters in a network.
Largest Betweenness	The largest number of shortest paths pass through all vertices

between the two groups. K-S is statistical testing that compares the distributions of two data samples to test whether they are from the same distribution. To be more specific, for each network metric, if misinformation sample tweets have cumulative distribution function (c.d.f) F_{mis}, and the true information samples have the c.d.f F_{true}, then using K-S we can test

$$H_0 : F_{mis} = F_{true}$$

versus

$$H_1 : F_{mis} \neq F_{true}$$

These statistical analyses reveal the difference of information dissemination between the two groups in terms of network aspects.

3.4.3 Results

To perform my analysis, I focused on the most popular tweets. One criterion to measure the popularity of a tweet is the number of times it is retweeted by retweeters other than its original poster. I assume a tweet to be popular if it is retweeted at least

50 times. About 5000 number of tweets in my dataset has retweets above this cut off. Among the top 5,000 most retweeted Zika tweets in 2016, a total of 400 tweets were identified and verified that contained misinformation. Among them, 266 tweets included adequate metadata to reconstruct the information dissemination networks. Not all metadata were available due to data loss, including ID banned by Twitter, content removed by Twitter, or user actively retracted the original post for various reasons. The comparison group of real information contained a total of 458 tweets that occurred within similar dates of posting of misinformation tweets and had a similar number of retweets of misinformation. To avoid potential selection bias, I did not make a one-to-one match of real Zika tweets.

3.4.3.1 Temporal Variability in Information Dissemination Dynamics

To even more examine the virality of tweets, I considered four more quantities. For every tweet, T_{50}, T_{75}, T_{90}, and T_{100} were defined as the time periods that the number of retweets reached to 50%, 75%, 90%, and 100% the total number of retweets of the tweet.

There was substantial temporal variability in the retweeting dynamics between misinformation and real information groups (Fig. 3.1). As it can be seen in Fig. 3.1, it took significantly less amount of time for misinformation to receive 50% of all retweets ($T_{50} = 334$ minutes for misinformation, $T_{50} = 448$ minutes for real information, $P < 0.001$ according to the two-sided t-test). The difference was minimal to receive 75% of all retweets ($T_{75} = 916$ minutes for misinformation, $T_{75} = 898$ minutes for real information, $P = 0.93$). Interestingly, it then always took significantly longer time for misinformation to receive 90% retweets ($T_{90} = 2580$ minutes vs $T_{90} = 1795$ minutes, $P = 0.03$), 95% retweets ($T_{95} = 4739$ minutes vs $T_{95} = 2824$ minutes, $P = 0.001$), and all retweets ($T_{100} = 34869$ minutes vs $T_{100} = 22340$ minutes, $P < 0.001$). These findings suggested that misinformation attracted or generated at least half of all retweets within a relatively short period of time in order to make

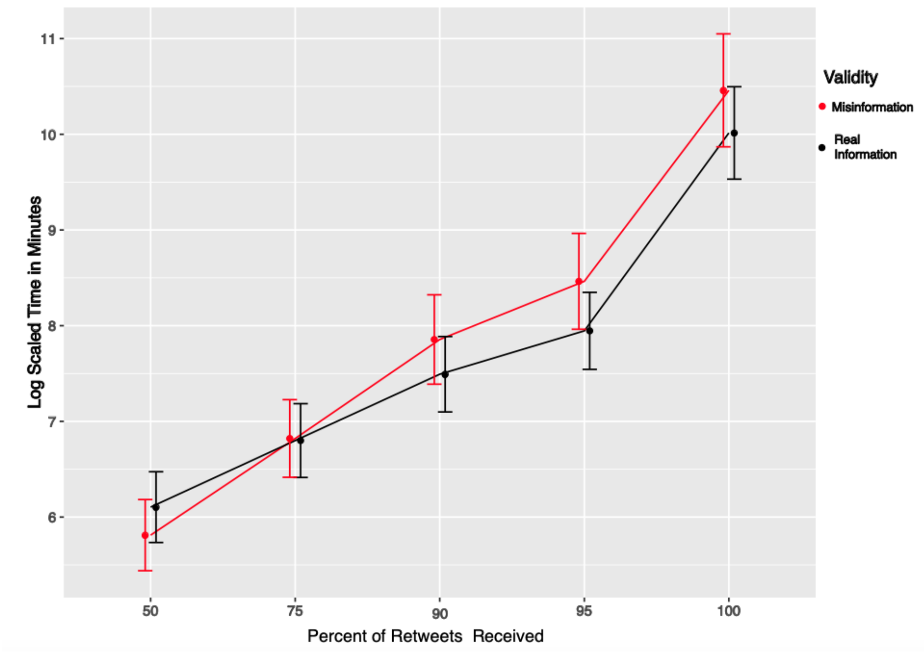


Figure 3.1: **Temporal Heterogeneity in Retweeting Activity in Zika Real and Misinformation Groups**. Y-axis (time) is in natural logarithm scale. X-axis (percent of retweets received) is not positioned by their absolute numbers. On average, misinformation (red) needed much shorter time to receive 50% of all its retweets comparing to real information (black), but significantly longer time to achieve 90% and above.

the misinformation more viral. Afterwards, misinformation might be deliberately retweeted to keep their visibility over a longer time span. Based on these observed temporal dynamics between the two groups, I chose the time till the last retweet to reconstruct the network, because it provided the most complete view of retweeting activity.

3.5 Differences of Network Metrics between Real and Misinformation Groups

I inferred and reconstructed the retweeting networks for each tweet in real and misinformation groups. Examples of dynamic network structures were provided in Fig. 3.3 for both misinformation and real information at different time points.

The distributions of important network metrics of both groups were computed and contrasted in Fig. 3.2. For demonstration purpose, all network metrics were scaled between 0 and 1 with feature scaling. Actual numeric summary statistics were shown in Table 3.3.

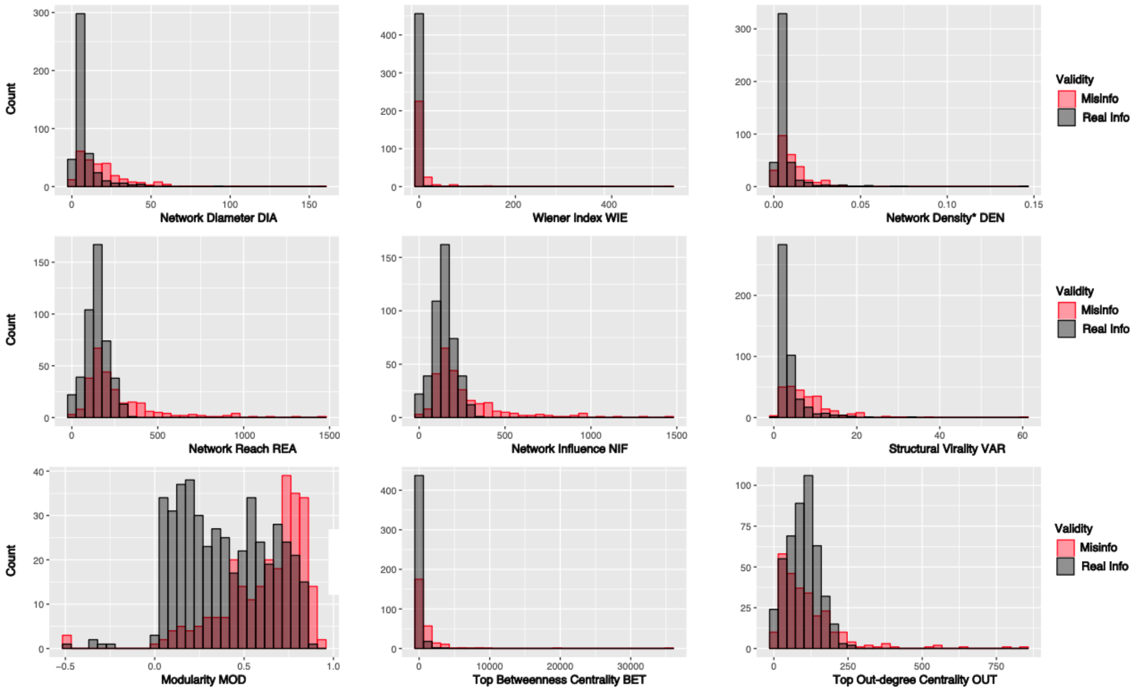


Figure 3.2: Distribution of Network Metrics

None of these distributions approximated normal distribution, showing high skewness and kurtosis as well as possible multimodality. This indicated large within-group variability of network structures. For any of these critical network metrics of information dissemination, the distributions always differed significantly ($p < 0.05$) between real and misinformation groups, according to Kolmogorov-Smirnov tests. Therefore, real and misinformation networks had a lot of heterogeneities, both within and between these groups.

Network density (DEN) was significantly higher in the misinformation group suggesting retweeters in the misinformation group were more likely to engage in retweeting misinformation if their friends (who they followed) tweeted or retweeted so. Note DEN was calculated on the initial diffusion network, which I constructed based on the follower-followee relationships of retweeters and the timestamp of their retweets, rather than the *information dissemination network*. The difference and relationship

Table 3.3: Statistics of Network Metrics in Zika Mis- and Real information Groups

	Misinformation			Real Information			
Network Metrics	Mean	S.E.	Median	Mean	S.E.	Median	K-S Test
Diameter	19.93	4.72	15	7.62	0.41	5	P<0.001
Wiener Index	884187	225669	113813	58013	4834	27067	P<0.001
Density	0.009	0.002	0.005	0.01	0.001	0.007	P<0.001
Size	264	14	190	145	3	142	P<0.001
Reach	263	14	188	144	2.99	141	P<0.001
Virality	8.01	0.4	6.49	3.82	0.15	2.62	P<0.001
Modularity	0.61	0.01	0.68	0.39	0.01	0.37	P<0.001
Betweenness	1003	170	278	127	15	15	P<0.001
Outdegree	115	6.88	89	97	2.43	100	P<0.001

between these two networks, and also the reason why I used the initial diffusion network to calculate density is explained in the previous section.

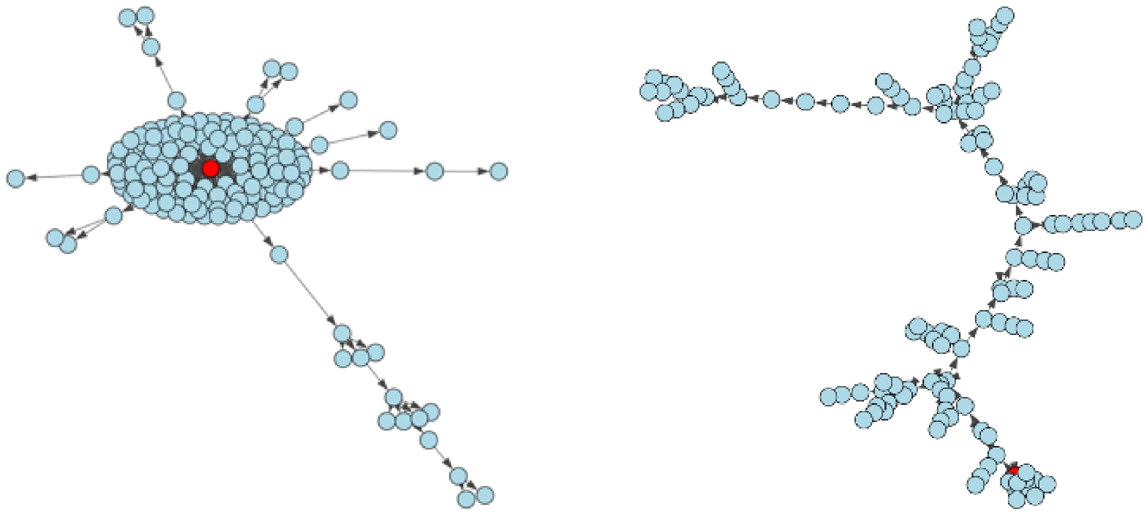
Network diameter (DIA) was also significantly higher in misinformation group. In general, the smaller diameter, the fewer layers the information (tweet) passed through the dissemination network to reach the outermost retweeters. My speculation was that Zika misinformation tweets attracted more grass-root users to retweet one after another, and form small clusters as opposed to a more star-like information dissemination network, in which most retweeters are directly connected to the original tweeter, in real information group.

Structural virality (VIR), which focused on average path length, was also significantly higher in the misinformation group, indicating vertices were generally further apart in the network. This finding confirmed our previous speculation that misinformation involved more direct user-to-user, or small cluster-to-cluster information dissemination than dissemination through fewer layers in real information group.

Network influence (NIF) and reach (REA) were similar metrics where REA specifically focused on unique retweeter. Zika misinformation group had both significantly smaller REA and NIF. In addition, I found that for Zika misinformation, about 30% of the network had the same vertex retweeted at least twice, which was substantially higher than in real information group at <10%. This could be an intentional propagation strategy to disseminate (mis)information on social media. However, the risk of such a strategy was that once other users noticed and reported this abnormal behavior, Twitter might take action to remove the suspicious tweet content and even ban the involved IDs. Therefore, having multiple IDs to retweet the same content together would be a more effective way to disseminate the information than to have the same ID to retweet the same content multiple times.

For the Wiener index (WIE), the misinformation group had significantly larger values on average. This indicated that misinformation retweeting networks had more chain-like form in comparison with true information networks that were more star-like and therefore had smaller WIEs. This finding also confirmed the previous finding that on average, network diameter (DIA) was larger in misinformation group, where networks were more chain-like with small local clusters as opposed to real information where retweeters were mainly connected to the original poster of the tweet and therefore forming a star.

For the misinformation group, WIE distribution had more than one prominent peak, i.e., multimodal. While some Zika misinformation networks had smaller WIE values, quite a few others had much larger WIE values (Fig. 3.2). From actual (mis)information propagation perspective, this implied that propagators exploited two seemingly contrasting strategies: the first one was to use a star-like network with very small WIE value (much less frequently observed in real information group), and the other one being chain-like dissemination network which had significantly large WIE value. In addition, there were hybrids of these strategies to disseminate misinformation further out. For example, propagators of misinformation might use bots or cyborgs to create an initial burst of retweeting activity, shown as local stars in the network, which attracted more actual grass-root human users to help pick up the trend and retweet one after another. However, I did not observe such a sophisticated arrangement in



(a) True information diffusion

(b) Misinformation diffusion

Figure 3.3: Diffusion networks of tweets

networks of real Zika information. Fig 3.3 shows examples of information dissemination networks from each misinformation and relevant group.

At the local network level, higher modularity was seen more frequently in the misinformation group, indicating that the user who retweeted misinformation tended to form smaller, local clusters to help disseminate the information. Therefore, Zika misinformation was more difficult to tackle using traditional misinformation mitigation strategies. Multiple smaller clusters reduced the risk of removal of some clusters, as other clusters served as alternative routes for information dissemination in the entire network. By comparison, the true information group had relatively smaller modularity. As can be seen in Fig. 3.2, MOD of misinformation group is more heavily skewed to the left compared to the real information group. In other words, in the misinformation group, MOD was dominated by upper bounds. This means that nodes in misinformation have more tendency to form clusters.

At individual vertex level, distribution of outdegree centrality OUT in misinformation group also had a strong multimodal pattern: this indicated that many misinformation tweets involved a user with an extremely large outbound degree (centrality score>200, Fig. 3.2) who might serve as potential online influential or propagator. On the other hand, the distribution of OUT in real information group was much similar to a normal distribution. OUT of misinformation group was more heavily right-skewed in comparison with the real information group. This shows the dominance of lower bounds in OUT, which in turn can be an indicator that a large number of users in the misinformation group could not be influential in relaying information.

For betweenness centrality (BET), top BET users in real information group had a significantly smaller BET score than that in the misinformation group (127 vs 1003). Therefore, these top BET users in misinformation group were more important than their counterparts in the real information group in terms of maintaining network stability, as a higher BET score indicated the more critical role of a vertex in relaying information. While top OUT users could be identified relatively easily by their superficial activity of attracting a large number of retweets, top BET users, on the other hand, were much more difficult to spot unless we were able to construct the network and perform centrality calculation for each vertex. Nevertheless, from a misinformation mitigation perspective, targeting top BET users could be an effective way to stall or even completely shut off misinformation propagation than focusing on top OUT users.

To summarize, the misinformation group had distinct distributions of all these network metrics from the real information group, indicating significantly different dissemination network structures. These findings from data mining of information dissemination networks could help health professionals and the general public better understand the dissemination pattern of health misinformation. In addition, these quantitative metrics could be utilized by health informaticians to develop more accurate health infosurveillance and misinformation detection systems.

3.6 Discussion

In this chapter, I developed an analytical framework to tackle the health misinformation dissemination problem on social media. I provided an operational definition of health misinformation and constructed an algorithm to explicitly track how health (mis)information disseminated on social media through retweeting networks. These novel discoveries provided solid shreds of evidence on the characteristics of misinformation dissemination patterns on Twitter, one of the most utilized social media platforms.

One of the key achievements of this study is to extract important features of health misinformation, which are not directly identifiable from the content of misinformation alone. I have shown the importance of treating misinformation (pathogen), users (hosts), and social media (environment) as an interconnected entity, the Infodemiology Triad. Misinformation, like real pathogens, are not leaving any trace behind. In my network analyses, I computed important network metrics for each tweets' dissemination network through retweeting. They substantially increased our understanding of misinformation dynamics on social media. Furthermore, the rich dataset can be used in conjunction with other features of misinformation (e.g., content, linguistic, and ID-based) to build a comprehensive health misinformation detector on social media. In the next chapter, I will use state-of-the-art machine learning methods to build comprehensive health misinformation classifiers.

I need to point out that our current knowledge about health topics evolve through time as more and more clinical, epidemiological, and other pieces of evidence become available, hence the idea "evidence-based". The term "real information" and "misinformation" should be used with caution because our current understanding might be falsified in the future. Timing of discussion should be considered, especially during an emerging health crises such as the Zika epidemic, such that our understanding of the epidemic refreshed rapidly. For example, I found that the Economist, a once considered reliable and reputable source of information, tweeted in Dec. 2016 that Zika is harmless to adults... (the post was now deleted) when at that time there had already been clear evidences [127] to show the causal effect of Zika virus infection and Guillain Barre syndrome (GBS) in adults. Had the tweet appeared in early 2016 when the causal relationship between Zika and GBS had not been established, it would not be deemed as misinformation. Consequently, an important follow-up of this study is to increase the health literacy in the society such that people learn how to check the validity of health information on social media, why it is misinformation, and frequently update their knowledge about the health issues, instead of merely being told whether a piece of information is real or not.

It is worth noting that in practice, to construct an information dissemination network from who-retweeted-whom social network using simply timestamp is not enough and other factors must be considered. To elaborate more, Twitter uses three modes to compose tweets within users' timelines; 1) tweets are ordered reversed-chronologically, 2) they are ranked based on several factors, such as recency, the tweets' author, and number of retweets or likes or 3) users can choose to see older tweets posted by accounts that they are generally engaged with ('In case you missed it'). Moreover, a user may find a tweet through search without being influenced by their friends. Therefore, a comprehensive framework to investigate the flow of information among retweeters should take all of these factors into consideration, establish a probability distribution upon them, and construct the information flow network based on that distribution. Devising such a framework is an interesting research direction but beyond the scope this paper. The ultimate goal of this study is to show that the flow of misinformation tweets is the different from relevant ones; thus, it is a key factor to be considered besides other characteristics of tweets (e.g., textual contents) to distinguish between them. Constructing a dissemination network based on only timestamp is not comprehensive but is totally valid since regardless how a retweeter has been exposed to a tweet before retweeting it, they could not be influenced by retweets after their own retweets. Using this logic, our proposed method helps to reduce the complexity of the initial follower-followee network to a more tractable one before any further analysis.

One follow-up research direction to our work is studying characteristics of users, especially identifying bots and examining their role in diffusion of each group of information. Identifying such user accounts was beyond the scope of this study; however, we examined a few user IDs manually and then we tested the suspicious ones with botometer, a bot detector tool proposed in [48], and available online [128]. The user IDs in the misinformation group received the higher score to be bot, although I should mention that bots are not necessarily malicious. As an example, a company might use bots to promote its new product on a social platform.

Out of 5,000 tweets that I examined, 5% were posted by verified users, mainly including news agencies, health organizations and also public figures. This observation agrees with other studies showing the strong presence of well-known news agencies and health organizations on Twitter during the Zika outbreak in 2016 [129]. In the misinformation group, the 'verified' user IDs mainly belonged to lobbying political media and organizations, and also individuals, including political figures and celebrities. This indicates that even non-malicious users could unintentionally get involved in relaying misinformation if they don't have enough domain knowledge to validate a piece of information. The non-verfied user_IDs in the misinformation group were individuals with mainly dubious names such as 'bugs bunny with rings and cigar' with handle 'trillballins', 1994 Subaru Outback with handle Sadieisonfire, and John with handle linnyitssn. On average, 0.012 of retweeters who participated in misinformation dissemination were verified. In the relevant group, the verified user_IDs were dominated by health organizations such as CDC and WHO and well-known news agencies, such as CNN, ABC News, and Fox News. The strong participation of these news agencies in the real group and their absence in the misinformation group shows that still these traditional media are highly reliable as a piece of news goes through different levels of fact-checking process before getting published by them. The non-verified user_IDs in the relevant group were mainly unknown users. On average, 0.02 of retweeters who participated in relaying real information were verified.

In the misinformation group, for the verified user_IDs the ratio of friends to followers was around 0.02 and for non-verified user_IDs it was around 0.4 on average. This quantity in the real group was about 0.0093 for the verified user IDs, and 0.5589 for non-verified user IDs. Overall, the ratio of friends to followers for well-known and verified users was small as they were followed by a large number of users but they followed fewer users in comparison.

Further work is currently underway to investigate users' (re)tweeting activity through time and how this temporal dynamics can also be utilized to explore misinformation infiltration. In this study, I constructed a static network G of a given tweet at the end of all retweeting activities. As I have shown the large temporal heterogeneity both within and between groups (Fig. 3.1), my developed algorithm is able to construct a dynamic network Gt. If a sudden rise in retweeting dynamics is detected at time t, a specific network ending at time t can be constructed to explicitly identify which retweeter is causing the burst of retweets, quantify the user's importance by calculating its centrality scores, and work at individual vertex level to further address the health misinformation epidemic on social media.

3.7 Conclusion

In this chapter, I investigated the problem of health-related misinformation propagation from a new perspective. I demonstrated that analyzing the dynamics of (mis)information dissemination among the users of a social platforms can provide salient features to distinguish misinformation from relevant information.

By investigating the structures of the networks of retweeters, and also the information flow tree, we concluded that Reach (REA), Influence (NIF), Diameter (DIA), Virality (VIR), Density (DEN), Wiener Index (WIE), Modularity (MOD), Out-Degree Centrality (OUT), and Betweenness Centrality (BET) are important attributes of information dissemination network to be considered to distinguish misinformation from relevant information.

In the next chapter, I investigate the strength of the proposed features to distinguish misinformation from relevant ones by building classification models on top of my discovered features and also previously discovered by other researchers. I believe that the valuable insights provided by this research work can significantly contribute to constructing more accurate and robust misinformation detection systems.

CHAPTER 4: Characterizing and Detecting Health-Related Misinformation on Social Media

4.1 Overview

In this chapter, I characterize health misinformation infiltration as a dynamic dissemination process on social media in addition to content-based features. Using the Zika discussion on Twitter in 2016 as the study system, I identified 264 most influential tweets with misinformation and matched 455 tweets with real information. I present the algorithm that I developed to infer the information dissemination network through the retweeting process of each tweet, and the extracted eight network metrics. I then demonstrate how information dissemination on Twitter can be approximated as a non-homogeneous Poisson process (NHPP) signal. Next, I propose the 40 signal features that I devised to characterize each NHPP. For content-based features, I discuss how I applied both LIWC and Doc2Vec to further extract 63 and 50 features for each tweet, respectively. I also considered 4 user features. Finally, I present the classification models which I trained based on all provided feature categories, and also two Machine Learning algorithms; Support Vector Machine (SVM) and Random Forest (RF) classifiers. Using all feature categories combined as input, an RF classifier achieved $> 83\%$ accuracy and $> 90\%$ AUC to detect misinformation.

4.2 Introduction

Social media platforms have provided public health professionals with valuable resources to study public opinions towards various health-related issues, and to use social media as one of the main outlets to efficiently and effectively spread accurate and timely information especially during health crisis such as Zika epidemic in 2016 and the current COVID-19 pandemic [97, 110].

Nevertheless, social media have also opened the room for misinformation and facilitated its infiltration and proliferation on the Internet during health emergencies [130]. Health misinformation is generally considered to be misleading, incorrect, not evidence-based, and malicious. These different types of health misinformation cause unnecessary confusion, anxiety, and anger of individuals rupture the society and seriously undermine health professionals' efforts in providing evidence-based knowledge and combating the real epidemic. WHO stated that it needed to fight two epidemics at the same time during the 2014 Ebola epidemic, one in the real world and the other on the Internet, especially on social media [131]. The 2016 Zika epidemic was another example where misinformation proliferated on social media, including Facebook and Twitter [97]. Unfortunately, increasing exposure to misinformation online and on social media can reinforce incorrect beliefs of users [132], making misinformation difficult to eradicate. Therefore, it is equally important, if not more, to reduce the influence of health misinformation on social media during health emergencies alongside curbing the epidemic itself. The first and foremost critical task is to accurately identify health misinformation on social media in order to neutralize them before they inflict harm to users.

Nevertheless, detecting health misinformation among the large volume of user-generated information on social media is a challenging task. State-of-the-art misinformation detection systems are generally based on fact-checking content of social media posts. There are several challenges associated with this approach. First, our knowledge of health issues increases rapidly, especially during an emerging health crisis. Such a fast pace of knowledge refreshing can make certain "knowledge" incorrect when new evidence emerges. For example, Zika was once assumed to transmit only through mosquito biting and from pregnant mother to fetus, but later evidence confirmed the possibility of sexual transmission as well. Second, fact-checking is only useful for objective contents but not effective against more subjective hate language, extreme bias, and discrimination against certain groups of people. Nevertheless, health, especially during large pandemics, is always confounded by complicated social, political, and economic issues. These convoluted issues make content-based fact-checking less useful against various types of misinformation. Third, content-based misinformation detection systems ignore other important features of health misinformation. Similar to a real epidemic, the digital epidemic of misinformation is a multi-aspect problem. The real pathogen, host, and environment together form the epidemiological triad, the foundation of an epidemic. We can tackle real epidemics more effectively when combining the power of etiology, epidemiology, immunology, and pathophysiology from this triad. Similarly, health misinformation, general users, and the social media environment also collectively form the infodemiological triad [133]. I suggest that a more holistic and accurate characterization of health misinformation is essential to develop a health misinformation detection system.

Recent studies have been exploring new features of health misinformation. Monitoring user activity on social media can detect suspicious online behavior and identify potential bots. While these content-based and user-based features are the key aspects to develop misinformation detection systems [134], other types of features of health misinformation have also been explored [135]. For instance, Qazvinian et al. constructed a probability distribution over retweeters of rumors and used it in addition to content-based features to predict rumors [136]. Another work investigated rumors on Sina Weibo, the largest social media platform in China, and proposed a combination of propagation-based, location-based, and client-based features [137]. Other studies investigate topic network features in addition to content-based features [38]. More recent studies also apply an array of machine learning (ML) and AI methods, time series analysis, and network analysis to build health misinformation detectors [138, 134], track misinformation dissemination dynamics [139, 140, 141], and evaluate the role of users (e.g., bots, trollers, opinion leaders) who relay misinformation [5, 142]. These studies provide alternative perspectives and technical approaches to identify health misinformation on social media.

In this study, I present a novel perspective in characterizing health misinformation on social media from multiple aspects. I further develop a data mining approach to extract new features that distinguish health misinformation from real ones, and develop more effective misinformation classifiers. First, I extract information dissemination features by constructing each tweet's retweeting networks to compute network metrics. In addition, I model the retweeting process of each tweet as a non-homogeneous Poisson process (NHPP) signal and extract signal-based features. Then I consider the content and linguistic features and apply LIWC and Doc2Vec. Finally, I extract user features (e.g., number of followers, friends, verification status) as well.

After extracting these different categories of features representing various aspects of health (mis)information, I develop classifiers using the most retweeted tweets containing real and misinformation about Zika in 2016. Different categories of features and their combinations are investigated as inputs to develop support vector machine (SVM) and random forest (RF) classifiers. I compare the performance of these models and evaluate the influence of input feature categories on classifiers' performance.

4.2.1 Data Retrieval

The Gnip API through UNC Charlotte School of Data Science was used to retrieve all English tweets with keyword *Zika* and other related keywords such as *microcephaly* and *PHEIC*. The entire year of 2016 (January 1 to December 31, 2016) was chosen as the sampling period for this study to bracket the entire WHO Public Health Emergency of International Concern (PHEIC) period from February 2 to November 18, 2016. This time period also covered major milestones in the Zika epidemic timeline, including WHO's initial warning of Zika epidemic across the Americas, the official PHEIC declaration of Zika pandemic on February 1, opening of Rio summer Olympics from August 5 to August 21, and the end of the PHEIC on November 18, 2016. A total of 3.7 million English tweets, retweets, and their associated metadata were collected in 2016. This dataset was the complete dataset covering all English discussions regarding Zika on Twitter in 2016, unlike the common 1% sampled data from Twitter's own API. Therefore, my dataset provided a more comprehensive, complete, and less biased view of Zika related tweets.

4.2.2 Tweet Annotation

All original Zika tweets were ranked based on the number of received retweets, from highest to lowest. I defined a tweet to be highly influential if received retweets from ≥ 50 distinct retweeters. Based on this criterion, the top 5,000 most retweeted Zika tweets were selected as the sample pool. An operational definition of misinformation regarding Zika was established, such that the content was not evidence-based, in accordance with the commonly used term evidence-based medicine in the health domain. Peer-reviewed journal articles and conference proceedings, government and health agencies (e.g., CDC and WHO) reports and statistics, fact-checking websites, are all used to evaluate and cross-check the tweets. Two independent researchers established strict intercoder reliability ($> 95\%$) on a randomly selected test set of 100 tweets before proceeding to annotate the remaining tweets. A total of 264 tweets were finally included as the misinformation group with complete metadata. Another 455 tweets were identified as real information group, controlling the effect of posting time and number of retweets to make them comparable to misinformation group. A detailed description of this annotation and misinformation identification process is provided in the previous chapter and in the study [143].

4.3 Features Extraction

4.3.1 Information Dissemination-Based Features

I extracted a novel set of features from these tweets and paved the way for further classification. In the first part of feature extraction, I considered (mis)information as a dynamic information dissemination process through retweeting. I explored two different but interrelated angles: 1) structural features of information dissemination network of retweeting, and 2) time series features of information dissemination signal. For the first feature category, I developed an algorithm to reconstruct and infer the retweeting network. For the second category, I constructed a signal of retweeting process and approximated it as a non-homogeneous Poisson process (NHPP) for each tweet. These two categories of features characterized (mis)information dissemination among users on social media.

4.3.1.1 Dynamic Retweeting Network Construction and Feature Extraction

I considered retweeting to occur in a network of retweeters for a given tweet. Such a retweeting network can be mathematically modeled as a graph, G, with two sets of elements: a set of nodes (or vertices, representing users), denoted by V, and a set of edges, denoted by E (representing retweeting sequence). $G = \{V, E\}$. Every edge e in E connects a pair of vertices v_i and v_j. In this study, pairs of vertices connected by edges were ordered (i.e., the direction of the edge is relevant), and the graph G was a directed graph. To construct an information dissemination network, I considered the followers' relationship among each retweeter and time difference of two consecutive retweets. For every pair of retweeters, (v_i, v_j), there was a directed edge e from v_j to v_i if and only if 1) v_j followed v_i, and 2) v_j had retweeted the tweet after v_i. A detailed description of this algorithm is provided in the previous chapter and also my prior work [143].

Once the retweeting network was constructed for each tweet, I further extracted critical network metrics for information dissemination in both real and misinformation groups. Network metrics between the two groups were compared by the Kolmogrov-Smirnov test to detect distribution differences between groups. I identified a total of nine network features with significant differences between real and misinformation groups. These nine features comprehensively characterize a dissemination network structure from global network level (density, reach, diameter, virality, influence, Wiener index) to local cluster level (modularity) and down to individual vertex level (top degree and top betweenness centrality score). The details of these network features and their relevance to information dissemination are in the previous chapter and also my prior work [143]. Evaluating the differences of these network features among real and misinformation groups provided insights on network dissemination differences, which is the cornerstone to build the classifier.

4.3.1.2 Retweeting Signal Construction and Feature Extraction

Network features characterized (mis)information dissemination among *users*. While I used the time difference of retweeting to infer dissemination network, the temporal aspect was not explicitly considered in the network features. Nevertheless, information dissemination, like a real epidemic, is a dynamic process where the temporal aspect is a critical factor. To more accurately capture the temporal dynamics of the retweeting process, I further constructed retweeting signals for each tweet, approximated them as non-homogeneous Poisson processes (NHPP), and extracted signal features thereof.

I constructed a time series of each tweet's retweeting stream, referred to as the retweeting signal hereafter. The signal can be mathematically described as $Y = Y_t, t \in T$ where Y_t is the number of retweets received during time t. To create a retweet signal for a tweet, first, the time window between its posting time and its last retweet was partitioned into fixed-size bins. Next, the number of retweets at each bin is counted, and a signal, Y_t is created out of the retweet counts. The option of bin size or temporal resolution is critical and would heavily influence any subsequent analysis. If the resolution is too high (e.g., at the second-based resolution by counting how many retweets occurred in each second), then one might be just modeling the noise. On the other hand, if the resolution is too low (e.g., at hourly resolution), then we might miss a lot of important signals, especially at the beginning of the retweet stream, when a lot of retweets usually are clustered. By carefully examining the retweet streams in my dataset, I concluded that for this study, the optimal temporal resolution could be obtained by 10-minute bins.

My initial exploration revealed a substantial temporal variability between real and misinformation groups[143]. It took significantly less amount of time for misinformation to receive 50% of all retweets but much longer time to receive 90%, 95%, and 100% retweets than real information group. I hypothesized that time to receive 50%, 75%, 90%, 95% retweets were useful to detect misinformation retweeting signal.

In addition, I detected peaks in retweeting signals to identify the time periods when a tweet was highly attractive and received a large number of retweets in a short period of time. Retweeting signals usually had more than one peaks, i.e., information relay. This complied with my previous findings that Twitter discussions could be impacted by real-world events and also promoted by influential users from time to time [110]. I found that time difference between two consecutive peaks p_i and p_{i+1} approximated exponential distribution. In other words, between every two successive peaks within a retweet signal, the number of retweets could be considered as a counting process that follows a Poisson distribution. Therefore, I modeled retweeting signal between two consecutive peaks as a specific homogeneous Poisson process (HPP), and the entire retweeting signal as a non-homogeneous Poisson process (*NHPP*) approximated by the union of n underlying $HPPs$:

$$NHPP = \bigcup_{i=1}^{n} HPP_i$$

Modeling a retweet process as a *NHPP*, in which the rate of events occurrence is not a fixed constant rate but a function of time t, correctly reflects my observation that it is common for multiple peaks to exist in retweeting signals.

NHPP is mathematically formulated as follows:

$$N(0) = 0, \tag{4.1}$$

$$P\{N(t+h) - N(t) = 1\} = \lambda(t)h + o(h), \tag{4.2}$$

$$P\{N(t+h) - N(t) \geq 2\} = o(h), \tag{4.3}$$

where $N(t+h) - N(t)$ is a Poisson random variable with the Poisson parameter λ as $E[N(t+h) - N(t)] = \int_t^{t+h} \lambda(s)ds$. At the beginning of the process, when the tweet is posted, the number of retweets is zero; $N(0) = 0$. The probability of receiving a retweet between time t and $t+h$ depends on the length h and the process rate at time t. In addition, the probability of occurring more than one retweet in the small time interval h is negligible.

Each HPP_i started at i-th peak, p_i, and continued until immediately before the next peak, when the next homogeneous Poisson process, HPP_{i+1} started. I characterized several important aspects of HPP_i: the peak,p_i, the valley, v_i, indicating the end of the process, and the HPP_i rate λ_i. I defined the valley v_i to be the time of the minimum number of retweets farthest from the peak p_i and closest to the p_{i+1} where the next HPP_{i+1} started. According to my exploration, a retweeting signal always had at most five peaks. Therefore, to make it consistent across all retweeting signals, I set, n, the maximum of the number of $HPPs$ for a signal, as 5. For each retweeting signal, I extracted the following features of its underlying HPP_i process:

- peak time p_i: the time when HPP_i started,
- peak height ph_i: number of retweet counts at peak i,

- Full width half max ($fwhm_i$): the first time point after each peak when the number of retweets was less than or equal to the half of the peak height,
- height of full width half max $fwhmh_i$: the number of retweet counts at the $i-th$ $FWHM$,
- valley height vh_i: the number of retweet counts at the $i-th$ valley,
- process width pw_i: the time difference between the peak p_i and valley v_i,
- valley time v_i: the time when HPP_i ended,
- λ_i: the HPP_i rate calculated as the average of retweet counts between p_i and v_i.

By extracting and comparing signal features, I obtained a more quantitative characterization of the temporal aspect of (mis)information dissemination on social media.

4.3.2 Content-Based Features

In addition to information dissemination feature categories, I explored and extracted various content-based features. Contents in misinformation, similar to the actual pathogens in a real transmission chain, interacted with human users (hosts) and exerted the influence psychologically. In this study, I focused on textual content and excluded other contents such as pictures, memes, and GIFs. Two major categories of content features, LIWC and Doc2Vec, were extracted to more comprehensively characterize content's topic, semantic and linguistic aspects.

4.3.2.1 LIWC: Linguistic Inquiry and Word Count Feature Extraction

To extract socially and psychologically relevant features from tweet content, I applied the widely used Linguistic Inquiry and Word Count (LIWC) tool [144]. LIWC summarized the tweet in various categories of emotional and cognitive patterns. Categories were represented and quantified by linguistic features of word counts and statistics. Examples of such features included positive or negative emotions, social relationships, social coordination, and individual differences. These features were more subjective for characterizing tweet content. According to my exploration of most retweeted Zika tweets, emotions (e.g., anger, confusion, fear, and mistrust) and social issues (e.g., bias and discrimination towards certain demographic groups) were profound along with this emerging disease and its discussion on social media.

After applying LIWC, each tweet received a numeric vector of 63 features using LIWC software. These features were further compared between real and misinformation groups. Such insights would be useful for a classifier to differentiate real and misinformation groups.

4.3.2.2 Doc2Vec: Document to Vector Feature Extraction

LIWC provided a quantitative way to characterize, extract, and quantify linguistic features. While LIWC vastly expanded my understanding of computational linguistics and paved the way for feature extraction, it is more subjective (e.g., related to social and psychological aspects), relied on a human-developed dictionary, and allowed for little flexibility of incorporating new insights especially on emerging health issues.

Word embedding techniques, on the other hand, provided a more direct and sophisticated representation, including semantics and relationships among words within the content. Word embedding used neural networks (NN) to learn and represent features of content. Each word is mapped into a pre-defined vector space where the values in the vector were learned from NN. This ML technique allowed for more natural capturing and representation of the content, comparing to LIWC. While word embedding also used a corpus (dictionary) to generate vector space, the corpus size was much larger than the one in LIWC and did not have human-assigned labels, thus allowing much more detailed and objective characterization of the content.

However, word embedding techniques had some shortcomings as well. They were not effective when analyzing short texts with different lengths, as in the case of tweets [145]. In this study, I applied an extension of the original word embedding technique, document-to-vector (doc2vec) proposed by Le et al. in [145]. In this new method, documents, regardless of their length were embedded into a pre-defined vector space. I treated each tweet as a document and applied doc2vec. After the initial evaluation and preprocessing (e.g., dropping non-text components such as figures, memes, and videos), I converted each input tweet to a document vector with a size of 50, i.e., 50 features in Doc2Vec vector space. These features could be explicitly compared between real and misinformation groups for further classification tasks.

4.3.3 User Feature Extraction

We extracted user features from their profiles, including number of followers, number of friends, verification status, and percentage of retweeters with verified status. These features were not associated with a specific tweet but to the user ID. Studies have used user features to detect potential malicious IDs which frequently sent out misinformation. However, I suggested that this approach could have both false negative and false positive issues. My more intensive exploration had identified some IDs which frequently sent out misinformation (e.g., @naturalnews, a far-right conspiracy theorist account); however, not all tweets it generated were misinformation. On the other side, some seemingly credible sources (e.g., @theEconomist) also sent out inaccurate information due to a lack of understanding of the evolving health issue [110]. I used user features in conjunction with tweet-specific features to provide a more comprehensive characterization of health misinformation.

4.3.4 Summary of Feature categories

In summary, I have identified and extracted five (5) categories of features: 9 dynamic network features and 40 signal features related to the information dissemination process through retweeting; 63 LIWC and 50 Doc2Vec features based on the actual content of each tweet. In addition, the fifth group of 4 user-based features is also extracted. Eventually, I extract 162 features for each tweet, both real and misinformation in this study, and perform further ML classification based on these categories of features.

4.4 Classification based on Multiple Feature Categories

I further built two types of supervised ML classifiers, random forest (RF) and support vector machine (SVM). I defined confirmed misinformation as *positive* while real information as *negative* in this study. Though there are many other classifiers available, I suggested the merit of this study was to provide a more comprehensive characterization of the health misinformation challenge on social media, and a data mining approach to extract new features. Therefore, the main goal of this study was not to develop new classifiers nor to systematically explore different classifiers' performance, but to develop effective classifier based on commonly used and proven robust algorithms such as RF and SVM.

For each type of classifiers (RF and SVM), I first included features from a single feature category. Then I built another set of classifiers by merging different feature categories. I ran k-fold cross-validation for each classifier I built. The data were split into 10 folds ($k = 10$) and randomly picked 9 slices to train the model while using the remaining unseen data to cross-validate the model. This process was repeated 10 times such that each slice was used exactly once for cross-validation.

4.4.1 Evaluation

To evaluate the constructed classification, I use the following metrics: accuracy, F1 score, AUC, defined as the area under the Receiver Operating Characteristic (ROC) Curve, the true positive rate (TPR), and the false positive rate (FPR), which are calculated and averaged from the 10-fold cross-validation process that explained before to evaluate RF and SVM classifiers' performance.

Accuracy is defined as the percentage of truly classified instances. The other mentioned metrics are derived from the confusion matrix which contains four elements: true positive (TP, classifier correctly identifies misinformation), false positive (FP, classifier incorrectly identifies real information as misinformation), true negative (TN, classifier correctly identifies real information), and false negative (FN, classifier incorrectly identifies misinformation as real information). In this study, the dataset was not balanced, and there were more tweets with real information than misinformation. Therefore, F1-score is a more representative performance metric than accuracy. F1-score is defined as the harmonic mean of precision and recall and calculated by

$$F1 = \frac{2 * precision * recall}{precision + recall};$$

where precision is calculated by

$$precision = \frac{TP}{TP + FP},$$

and recall (also called true positive rate,TPR) is calculated as

$$recall = \frac{TP}{TP + FN}.$$

ROC curve and AUC help visually compare two models performance. For a non-skill classifier, which is hardly able to distinguish between positives and negatives, AUC is close to 0.5 as the baseline. The better the classifier is, the closer its AUC gets to 1 as the asymptotic limit, which is the maximum performance a classifier can achieve. Data processing, mining, and exploratory analyses were carried in *R* 3.5.0 with various supporting packages. ML classifiers were developed in *Python* 3.7 with *SciKitLearn*. All annotated data and accompanying codes will be freely available on open data and code sharing repository (GitHub).

4.5 Results

4.5.1 Feature Comparison between Zika Real and Misinformation Tweets

First, I show how network features differ between Zika real and misinformation tweet groups in Fig. 4.1. Note that these features are scaled between 0 and 1, so they do not represent actual values in Fig. 4.1 through Fig. 4.4. All 9 network features, characterizing information dissemination network from the global level to individual vertex level, differ significantly ($p < 0.05$) between the two groups according to the Kolmogrov-Smirnov test. Based on RF classification results in the later section, the most influential network feature, as measured by Gini impurity, is the total path (TPA), followed by Wiener index (WIE) and virality (VIR). These results reveal information dissemination network structures differ substantially between real and misinformation groups.

For signal features, I identify the top 9 most influential features using Gini impurity out of a total of 40 features, making it consistent with network features. A comparison of these 9 important signal features is shown in Fig. 4.2. Interestingly, the height of the 3rd peak, not the first two peaks, is the strongest signal to differentiate misinformation from real information. My explanation is that most real information signals may not have more than 2 peaks, thus having a 3rd peak could be a sign of misinformation. The next important signal feature is time to receive 50% retweets, also known as the "half-life" of a tweet. The third important feature is the valley width of 1st peak, i.e., time duration between 1st and 2nd peaks. This feature quantifies the rate of information relay and can be used to distinguish misinformation. Kolmogrov-Smirnov tests further show that these 9 most influential signal features' distributions also differed significantly ($p < 0.05$) between real and misinformation groups.

In addition to information dissemination feature categories, I characterized Zika real and misinformation content features on Twitter. First, I identified a list of top differentiating LIWC features (Fig. 4.3), which were significantly different between the two groups ($p < 0.05$). "Money" was the most important feature, and misinformation was related to the monetary aspect of Zika much more frequently than real information. The next was "Analytic," where real information had a much higher likelihood to involve analytical aspects than misinformation. All these results showed content topic differences between Zika real and misinformation groups. It reinforced my suggestion that health is not an isolated issue but is confounded with various social, political, and economic issues, which may lead to potential misinformation.

Next, I showed Doc2Vec feature differences between Zika real and misinformation groups (Fig. 4.4). According to the later RF model, feature 24 was the single most influential feature, accounting for more than 75% of decision tree split based on Gini impurity and dwarfing all other Doc2Vec features. However, unlike LIWC features, Doc2Vec features were learned by NN directly and did not have a clear interpretation.

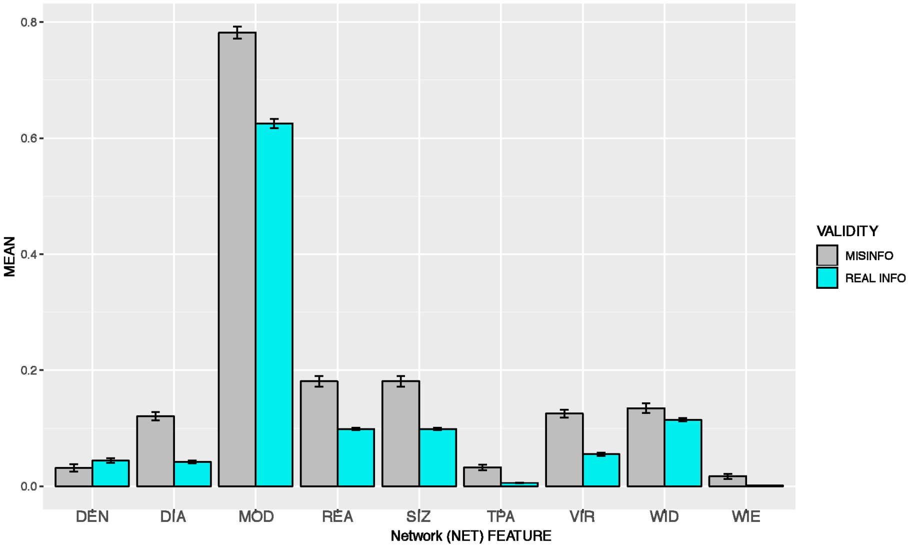


Figure 4.1: Network Features Comparison between Real and Misinformation Groups

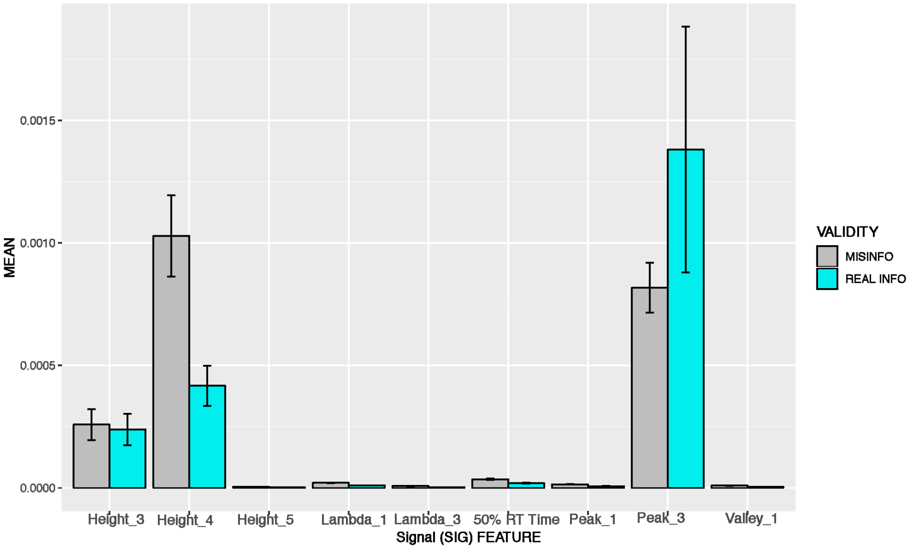


Figure 4.2: Top Signal Features Comparison between Real and Misinformation Groups

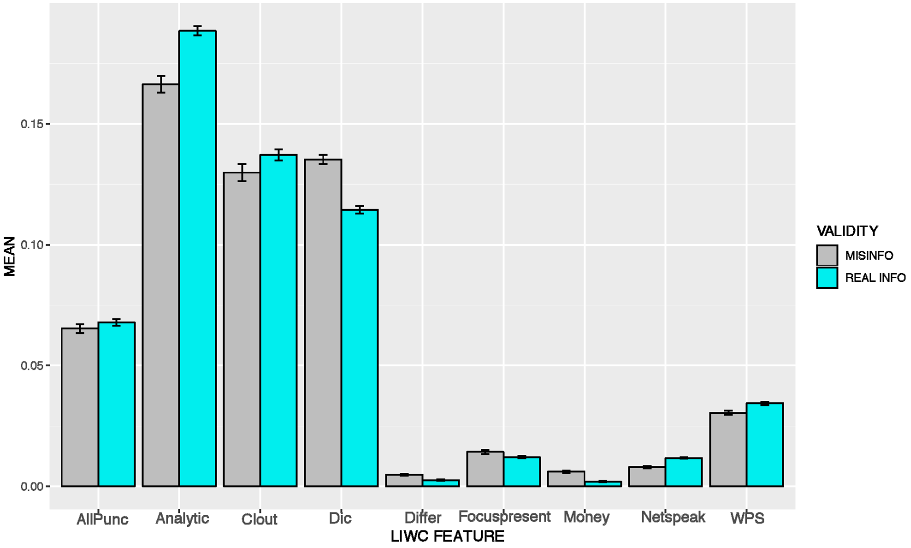


Figure 4.3: Top LIWC Features Comparison between Real and Misinformation Groups

4.5.2 Machine Learning Classification and Detection of Zika Misinformation

Table 4.1 and 4.2 show classifier performance of RF and SVM, respectively. In general, a single feature group as input is less effective in detecting Zika misinformation on Twitter. Among the five feature categories I have explored, content-based DOC2VEC (D2V) features show the best performance with RF model, followed by

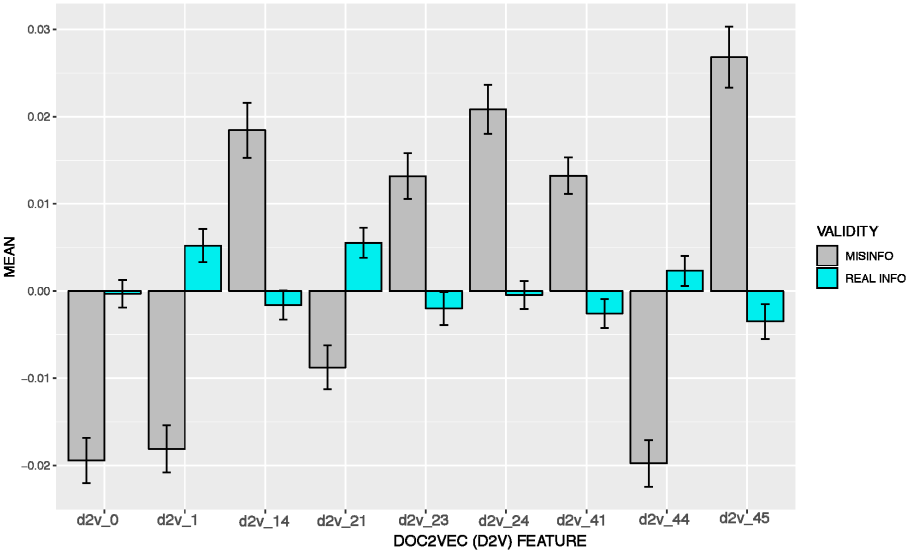


Figure 4.4: Top Doc2Vec Features Comparison between Real and Misinformation Groups

user-related (USER), LIWC, and network (NET) features. Signal (SIG) features alone have the least power to differentiate misinformation from real ones. Using SVM, the same results still hold where DOC2VEC is the most differentiating group of features. Combining different groups of features increase the classifier's performance. For RF, combining NET and SIG as dissemination features slightly increase model accuracy, F1 score, and AUC (73%, 79%, 78%) from NET (71%, 77%, 77%) and SIG (70%, 77%, 75%) group alone. Interestingly, combining LIWC and D2V as content features does not increase model performance. The highest model performance is achieved when combining all five categories of features, with accuracy, F1 score, and AUC at 82%, 86%, and 90%, respectively (4.1, last row). These results demonstrate that health misinformation on social media is a multi-aspect problem, and we need to characterize different aspects of health misinformation more comprehensively. The comparison of ROC curves and AUC values across combinations of features in RF is shown in Fig. 4.5.

Comparing between the two classifiers, RF always outperforms SVM in this study. Combining all feature categories still yields the best performance in SVM (4.2, last row). Nevertheless, the best-performed SVM classifier trails behind RF with accuracy, F1 score, and AUC at 77%, 82%, and 95%, respectively, with a -5%, -4%, and -5% difference from the best RF model. The comparison of ROC Curve and AUC values across combinations of features in SVM is shown in Fig. 4.6.

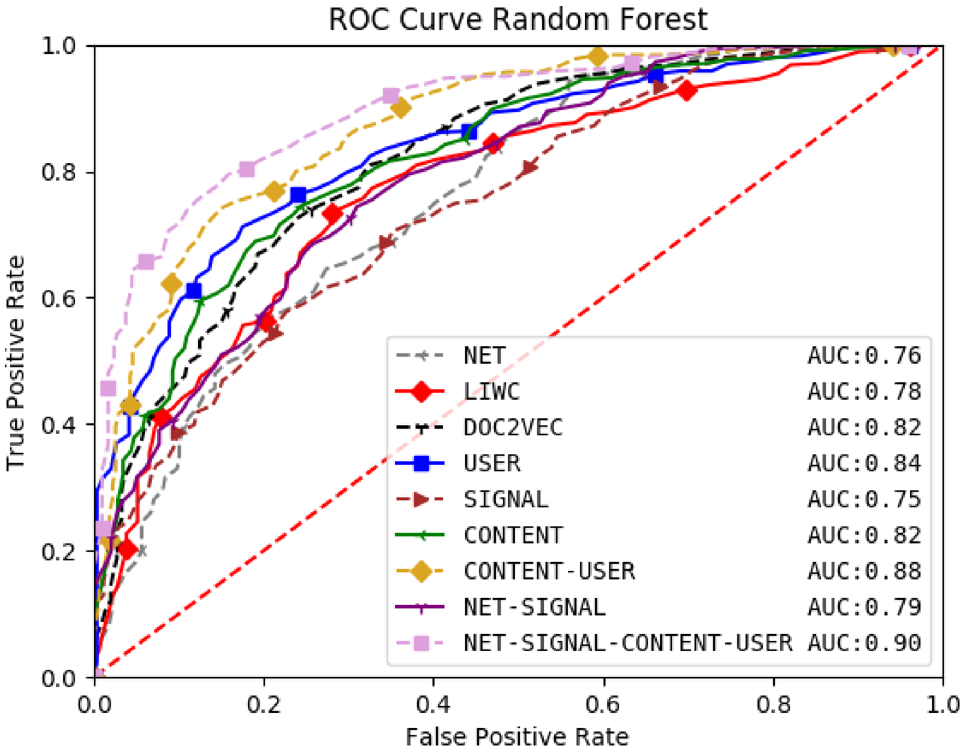


Figure 4.5: ROC and AUC of RF Classifier with Different Feature Categories

4.6 Discussion

This study delivers an accurate health misinformation classifier that takes multiple aspects of misinformation into consideration. More importantly, this study provides a more holistic view of the health misinformation challenge on social media by exploring

Table 4.1: Performance with Different Feature Categories: RF

Feature Categories	**Accuracy**	**F1-score**	**AUC**	**TPR**	**FPR**
USER	0.770	0.823	0.839	0.87	0.061
LIWC	0.731	0.791	0.776	0.83	0.056
DOC2VEC (D2V)	0.771	0.830	0.821	0.91	0.54
NETWORK (NET)	0.711	0.772	0.765	0.8	0.57
SIGNAL (SIG)	0.697	0.769	0.75	0.82	0.50
LIWC+D2V (Content)	0.750	0.807	0.825	0.85	0.058
NET+SIG	0.733	0.787	0.782	0.81	0.60
USER+LIWC+D2V	0.796	0.841	0.877	0.88	0.65
All Feature Categories	**0.822**	**0.859**	**0.901**	0.89	0.72

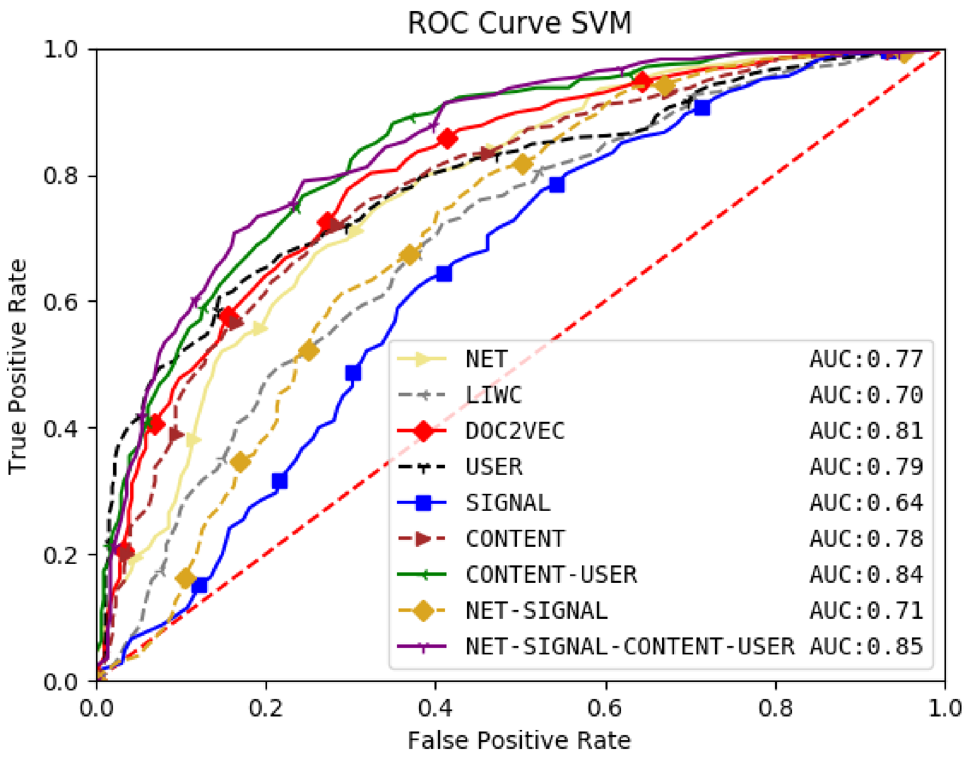


Figure 4.6: ROC and AUC of SVM Classifier with Different Feature Categories

different aspects of misinformation, not just the content itself. Health misinformation is like a pathogen in the real world, and no single discipline can tackle pandemic alone. Therefore, combining power from different aspects of health misinformation will enable better understanding and response to health misinformation challenge on

Table 4.2: Performance with Different Feature Categories: SVM

Feature Categories	**Accuracy**	**F1-score**	**AUC**	**TPR**	**FPR**
USER	0.679	0.765	0.792	0.85	0.41
LIWC	0.678	0.756	0.698	0.81	0.46
DOC2VEC (D2V)	0.753	0.810	0.807	0.87	0.56
NETWORK (NET)	0.737	0.808	0.774	0.90	0.48
SIGNAL (SIG)	0.669	0.761	0.645	0.86	0.35
LIWC+D2V (Content)	0.729	0.787	0.782	0.81	0.60
NET+SIG	0.707	0.787	0.782	0.81	0.60
USER+LIWC+D2V	0.768	0.814	0.843	0.83	0.68
All Feature Categories	**0.771**	**0.817**	**0.847**	0.83	0.68

social media.

Practically, this study utilizes data mining techniques to extract more features, especially information dissemination features. The retweeting network $G = \{V, E\}$ in this study is constructed at the end of information dissemination when the last retweet is received. Future work can be done to further construct and characterize temporal network $G_t = \{V, E, t\}$ at given time t, and merge insights with signal-based features that explicitly tackle the temporal aspect of (mis)information dissemination.

Findings from this study not only deliver an accurate Zika misinformation classifier on social media, but also shed new insights on characterizing and understanding general misinformation, including health misinformation more comprehensively on social media. The new perspective and approach can be readily transferred to tackle other emerging issues such as the current COVID-19 pandemic, and to develop more generic classifiers.

However, I need to point out some limitations of this work. First, the feature categories are not an exhaustive list, and new features are yet to be discovered. Like pathogens in the real world, misinformation can also adapt to the changing environment, mimic behavior of real information, and become more difficult to detect. Second, this study emphasizes more on feature extraction, and future work will continue identifying the most effective set of input features across feature categories. I have applied RF to rank feature importance based on Gini impurity, and this will guide further fine-tuning of ML classifiers with fewer inputs. In addition, I will evaluate the potential over-fitting issue in delivering ML classifiers. Third, this classifier is developed in the context of the 2016 Zika discussion, and its effectiveness needs to be re-evaluated in more recent health issues such as current COVID-19 where the social media landscape has changed since 2016. I will apply transfer learning techniques to adopt this work in new health emergencies.

4.7 Conclusion

In this study, I transfer existing knowledge of real epidemics to comprehensively characterize Zika misinformation infiltration on social media in 2016. I develop a novel data mining technique to construct (mis)information dissemination networks and signals, extract dissemination features, and further combine content-based features based on LIWC and Doc2Vec and user features. Based on the combinations of different feature categories, I develop an accurate Zika misinformation classifier that can detect misinformation with $> 85\%$ accuracy and $> 90\%$ AUC. The novel perspective and analytical framework in this study can be transferred to respond to misinformation during current COVID-19 and future pandemics.

CHAPTER 5: Future Works and Conclusion

5.1 Future Works

My doctoral research can be further extended in a number of ways. One direction is to devise frameworks that can help researchers better understand detected trends by summarizing the underlying discussions forming the trends, identifying the conflicting ones, clustering, and characterizing the sub-populations that are initiating such discussions. Another interesting direction is to construct multilingual frameworks that can aggregate user-generated content in different languages, which can be highly beneficial to improve risk communication during a worldwide crisis such as COVID-19. Another possibility is studying the role of users in distributing information on a social media platform. On the one hand, to contain and prevent misinformation propagation is by identifying malicious users who trigger and promote misleading information. On the other hand, to promote and facilitate the propagation of true information, we can rely on benign influential users within a network.

The proposed misinformation detection system in chapter 4 can be further improved and extended as a more versatile tool by rigorously testing more datasets from various domains, such as politics, and also in the context of other social media platforms, such as Facebook. One way to achieve this is by participating in or even organizing competitions in platforms such as Kaggle [146]. At the time of writing this manuscript, I found 12 competitions related to fake news detection [147], out of which only one had network-related data [148]. One research work based on the dataset [149] of this competition is [150]. Although Zhou et al. in [150] also proposed a set of network-based features, these features are extracted from a different type of network. They concentrate on a topic (i.e., a news article extracted from Buzzfeed or Politifacts) and construct a network of friendship from the users who engaged with that topic. They assume engagement to be posting a tweet, retweeting, liking, or replying. Next, based on some psychological theories (e.g., "More users spread fake news than true news"), they compute susceptibility scores for every node (user) in a network. These susceptibility scores, which are the main foundation of network-based features proposed in [150], are calculated based on the history of users in engaging with previous fake news. This can cause some limitations as historical data might not be readily available.

5.2 Conclusion

Social media platforms can play a critical role in the public health domain by providing a valuable resource to study the public's opinion regarding health-related issues. Moreover, they can be used as a powerful tool to relay beneficial messages to boost public knowledge of health-related topics.

In this doctoral dissertation, I designed and developed different computational models to analyze health-related information dissemination on Twitter from various aspects. In chapter 2, I investigated the driving factors that impact health-related discussions. I proposed and examined two hypotheses about the potential driving factors; 1) real-world events, and 2) highly active and also influential users. In particular, I designed and developed a computational pipeline, EventPeriscope, to measure the reflection of real-world events on Twitter discussions. This pipeline uses signal processing techniques to model tweet streams and to detect peaks showing the time points when users were highly active. Next, using text mining and NLP techniques, it digs into the textual contents of tweets at the peak time to measure the relevance of tweets to the topic of interest. Eventperiscope can help analysts to better understand the users' engagement with a specific health-related topic. In chapter 3, I proposed a framework to model information diffusion on Twitter. By analyzing the constructed information diffusion networks, I demonstrate the misinformation is propagated differently from true information. In chapter 4, I proposed a solution to model the retweeting process of tweets as NHPPs. By analyzing these processes and also the findings from chapter 3, I devised a new system that uses various aspects of information propagation (e.g., content, propagators, and temporal patterns) to detect misinformation with a high level of accuracy.

REFERENCES

[1] D. M. Lazer, M. A. Baum, Y. Benkler, A. J. Berinsky, K. M. Greenhill, F. Menczer, M. J. Metzger, B. Nyhan, G. Pennycook, D. Rothschild, *et al.*, “The science of fake news,” *Science*, vol. 359, no. 6380, pp. 1094–1096, 2018.

[2] H. Allcott and M. Gentzkow, “Social media and fake news in the 2016 election,” *Journal of economic perspectives*, vol. 31, no. 2, pp. 211–36, 2017.

[3] N. Grinberg, K. Joseph, L. Friedland, B. Swire-Thompson, and D. Lazer, “Fake news on twitter during the 2016 us presidential election,” *Science*, vol. 363, no. 6425, pp. 374–378, 2019.

[4] L. Z. Cooper, H. J. Larson, and S. L. Katz, “Protecting public trust in immunization,” *Pediatrics*, vol. 122, no. 1, pp. 149–153, 2008.

[5] D. A. Broniatowski, A. M. Jamison, S. Qi, L. AlKulaib, T. Chen, A. Benton, S. C. Quinn, and M. Dredze, “Weaponized health communication: Twitter bots and russian trolls amplify the vaccine debate,” *American journal of public health*, vol. 108, no. 10, pp. 1378–1384, 2018.

[6] M. Sharma, K. Yadav, N. Yadav, and K. C. Ferdinand, “Zika virus pandemicâ-analysis of facebook as a social media health information platform,” *American journal of infection control*, vol. 45, no. 3, pp. 301–302, 2017.

[7] L. Bode and E. K. Vraga, “See something, say something: Correction of global health misinformation on social media,” *Health communication*, vol. 33, no. 9, pp. 1131–1140, 2018.

[8] P. Sharma and P. D. Kaur, “Effectiveness of web-based social sensing in health information disseminationâa review,” *Telematics and Informatics*, vol. 34, no. 1, pp. 194–219, 2017.

[9] E. Z. Kontos, K. M. Emmons, E. Puleo, and K. Viswanath, “Communication inequalities and public health implications of adult social networking site use in the united states,” *Journal of health communication*, vol. 15, no. sup3, pp. 216–235, 2010.

[10] J. C. Eichstaedt, R. J. Smith, R. M. Merchant, L. H. Ungar, P. Crutchley, D. Preoţiuc-Pietro, D. A. Asch, and H. A. Schwartz, “Facebook language predicts depression in medical records,” *Proceedings of the National Academy of Sciences*, vol. 115, no. 44, pp. 11203–11208, 2018.

[11] B. Curtis, S. Giorgi, A. E. Buffone, L. H. Ungar, R. D. Ashford, J. Hemmons, D. Summers, C. Hamilton, and H. A. Schwartz, “Can twitter be used to predict county excessive alcohol consumption rates?,” *PloS one*, vol. 13, no. 4, p. e0194290, 2018.

[12] C. St Louis and G. Zorlu, "Can twitter predict disease outbreaks?," *Bmj*, vol. 344, p. e2353, 2012.

[13] C. W. Schmidt, "Trending now: using social media to predict and track disease outbreaks," 2012.

[14] J. P. Alperin, C. J. Gomez, and S. Haustein, "Identifying diffusion patterns of research articles on twitter: A case study of online engagement with open access articles," *Public Understanding of Science*, vol. 28, no. 1, pp. 2–18, 2019.

[15] J. K.-S. Liew and G. Z. Wang, "Twitter sentiment and ipo performance: a cross-sectional examination," *Journal of Portfolio Management*, vol. 42, no. 4, p. 129, 2016.

[16] N. Panagiotou, I. Katakis, and D. Gunopulos, "Detecting events in online social networks: Definitions, trends and challenges," in *Solving Large Scale Learning Tasks. Challenges and Algorithms*, pp. 42–84, Springer, 2016.

[17] M. Hasan, M. A. Orgun, and R. Schwitter, "A survey on real-time event detection from the twitter data stream," *Journal of Information Science*, p. 0165551517698564, 2017.

[18] M. Mathioudakis and N. Koudas, "Twittermonitor: trend detection over the twitter stream," in *Proceedings of the 2010 ACM SIGMOD International Conference on Management of data*, pp. 1155–1158, ACM, 2010.

[19] J. Weng and B.-S. Lee, "Event detection in twitter.," *ICWSM*, vol. 11, pp. 401–408, 2011.

[20] M. Cordeiro, "Twitter event detection: combining wavelet analysis and topic inference summarization," in *Doctoral symposium on informatics engineering*, pp. 11–16, 2012.

[21] W. Dou, X. Wang, D. Skau, W. Ribarsky, and M. X. Zhou, "Leadline: Interactive visual analysis of text data through event identification and exploration," in *Visual Analytics Science and Technology (VAST), 2012 IEEE Conference on*, pp. 93–102, IEEE, 2012.

[22] Y. Ren, R. Wang, and D. Ji, "A topic-enhanced word embedding for twitter sentiment classification," *Information Sciences*, vol. 369, pp. 188–198, 2016.

[23] S. B. Kaleel and A. Abhari, "Cluster-discovery of twitter messages for event detection and trending," *Journal of Computational Science*, vol. 6, pp. 47–57, 2015.

[24] M. Cha, H. Haddadi, F. Benevenuto, and K. P. Gummadi, "Measuring user influence in twitter: The million follower fallacy," in *fourth international AAAI conference on weblogs and social media*, 2010.

[25] P. Stamatiou, J. McCree, T. Marhsall, and M. Robertson, “Twitter. cs4803: Design of online communities,” 2008.

[26] W. W. Xu, Y. Sang, S. Blasiola, and H. W. Park, “Predicting opinion leaders in twitter activism networks: The case of the wisconsin recall election,” *American Behavioral Scientist*, vol. 58, no. 10, pp. 1278–1293, 2014.

[27] C. S. Park and B. K. Kaye, “The tweet goes on: Interconnection of twitter opinion leadership, network size, and civic engagement,” *Computers in Human Behavior*, vol. 69, pp. 174–180, 2017.

[28] B. Berelson, H. Gaudet, and P. F. Lazarsfeld, *The people’s choice: How the voter makes up his mind in a presidential campaign*. Columbia University Press, 1968.

[29] M. Vergeer, “Twitter and political campaigning,” *Sociology compass*, vol. 9, no. 9, pp. 745–760, 2015.

[30] S. Choi, “The two-step flow of communication in twitter-based public forums,” *Social Science Computer Review*, vol. 33, no. 6, pp. 696–711, 2015.

[31] Y. Hwang, “Does opinion leadership increase the followers on twitter,” *International Journal of Social Science and Humanity*, vol. 5, no. 3, p. 258, 2015.

[32] R. Karlsen, “Followers are opinion leaders: The role of people in the flow of political communication on and beyond social networking sites,” *European Journal of Communication*, vol. 30, no. 3, pp. 301–318, 2015.

[33] J. Riddell, A. Brown, I. Kovic, and J. Jauregui, “Who are the most influential emergency physicians on twitter?,” *Western Journal of Emergency Medicine*, vol. 18, no. 2, p. 281, 2017.

[34] J. Heidemann, M. Klier, and F. Probst, “Identifying key users in online social networks: A pagerank based approach,” 2010.

[35] A. Aleahmad, P. Karisani, M. Rahgozar, and F. Oroumchian, “Olfinder: Finding opinion leaders in online social networks,” *Journal of Information Science*, vol. 42, no. 5, pp. 659–674, 2016.

[36] V. V. Vydiswaran and M. Reddy, “Identifying peer experts in online health forums,” *BMC medical informatics and decision making*, vol. 19, no. 3, p. 68, 2019.

[37] S. Okazaki, A. M. Díaz-Martín, M. Rozano, and H. D. Menéndez-Benito, “Using twitter to engage with customers: a data mining approach,” *Internet Research*, vol. 25, no. 3, pp. 416–434, 2015.

[38] C. Castillo, M. Mendoza, and B. Poblete, “Information credibility on twitter,” in *Proceedings of the 20th international conference on World wide web*, pp. 675–684, ACM, 2011.

[39] J. ODonovan, B. Kang, G. Meyer, T. Höllerer, and S. Adalii, "Credibility in context: An analysis of feature distributions in twitter," in *2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Confernece on Social Computing*, pp. 293–301, IEEE, 2012.

[40] H. Allcott, M. Gentzkow, and C. Yu, "Trends in the diffusion of misinformation on social media," tech. rep., National Bureau of Economic Research, 2019.

[41] A. Gupta, P. Kumaraguru, C. Castillo, and P. Meier, "Tweetcred: Real-time credibility assessment of content on twitter," in *International Conference on Social Informatics*, pp. 228–243, Springer, 2014.

[42] P. Resnick, S. Carton, S. Park, Y. Shen, and N. Zeffer, "Rumorlens: A system for analyzing the impact of rumors and corrections in social media," in *Proc. Computational Journalism Conference*, pp. 10121–0701, 2014.

[43] P. T. Metaxas, S. Finn, and E. Mustafaraj, "Using twittertrails. com to investigate rumor propagation," in *Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing*, pp. 69–72, ACM, 2015.

[44] N. Hassan, A. Sultana, Y. Wu, G. Zhang, C. Li, J. Yang, and C. Yu, "Data in, fact out: automated monitoring of facts by factwatcher," *Proceedings of the VLDB Endowment*, vol. 7, no. 13, pp. 1557–1560, 2014.

[45] C. Shao, G. L. Ciampaglia, A. Flammini, and F. Menczer, "Hoaxy: A platform for tracking online misinformation," in *Proceedings of the 25th international conference companion on world wide web*, pp. 745–750, International World Wide Web Conferences Steering Committee, 2016.

[46] C. Shao, G. L. Ciampaglia, O. Varol, K.-C. Yang, A. Flammini, and F. Menczer, "The spread of low-credibility content by social bots," *Nature communications*, vol. 9, no. 1, p. 4787, 2018.

[47] J. Ma, W. Gao, P. Mitra, S. Kwon, B. J. Jansen, K.-F. Wong, and M. Cha, "Detecting rumors from microblogs with recurrent neural networks.," in *Ijcai*, pp. 3818–3824, 2016.

[48] O. Varol, E. Ferrara, C. A. Davis, F. Menczer, and A. Flammini, "Online human-bot interactions: Detection, estimation, and characterization," in *Eleventh international AAAI conference on web and social media*, 2017.

[49] Z. Zhao, P. Resnick, and Q. Mei, "Enquiring minds: Early detection of rumors in social media from enquiry posts," in *Proceedings of the 24th International Conference on World Wide Web*, pp. 1395–1405, International World Wide Web Conferences Steering Committee, 2015.

[50] L. Wu and H. Liu, "Tracing fake-news footprints: Characterizing social media messages by how they propagate," in *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, pp. 637–645, ACM, 2018.

[51] S. Vosoughi, D. Roy, and S. Aral, "The spread of true and false news online," *Science*, vol. 359, no. 6380, pp. 1146–1151, 2018.

[52] F. Alshaikh, F. Ramzan, S. Rawaf, and A. Majeed, "Social network sites as a mode to collect health data: a systematic review," *Journal of medical Internet research*, vol. 16, no. 7, p. e171, 2014.

[53] P. Balatsoukas, C. M. Kennedy, I. Buchan, J. Powell, and J. Ainsworth, "The role of social network technologies in online health promotion: a narrative review of theoretical and empirical factors influencing intervention effectiveness," *Journal of medical Internet research*, vol. 17, no. 6, p. e141, 2015.

[54] L. E. Charles-Smith, T. L. Reynolds, M. A. Cameron, M. Conway, E. H. Lau, J. M. Olsen, J. A. Pavlin, M. Shigematsu, L. C. Streichert, K. J. Suda, *et al.*, "Using social media for actionable disease surveillance and outbreak management: a systematic literature review," *PloS one*, vol. 10, no. 10, p. e0139701, 2015.

[55] F. J. Grajales III, S. Sheps, K. Ho, H. Novak-Lauscher, and G. Eysenbach, "Social media: a review and tutorial of applications in medicine and health care," *Journal of medical Internet research*, vol. 16, no. 2, p. e13, 2014.

[56] M. P. Hamm, A. Chisholm, J. Shulhan, A. Milne, S. D. Scott, L. M. Given, and L. Hartling, "Social media use among patients and caregivers: a scoping review," *BMJ open*, vol. 3, no. 5, p. e002819, 2013.

[57] Y. Hu, "Health communication research in the digital age: a systematic review," *Journal of Communication in Healthcare*, vol. 8, no. 4, pp. 260–288, 2015.

[58] T. A. Kass-Hout and H. Alhinnawi, "Social media in public health," *Br Med Bull*, vol. 108, no. 1, pp. 5–24, 2013.

[59] J. Lardon, R. Abdellaoui, F. Bellet, H. Asfari, J. Souvignet, N. Texier, M.-C. Jaulent, M.-N. Beyens, A. Burgun, and C. Bousquet, "Adverse drug reaction identification and extraction in social media: a scoping review," *Journal of medical Internet research*, vol. 17, no. 7, p. e171, 2015.

[60] C. A. Maher, L. K. Lewis, K. Ferrar, S. Marshall, I. De Bourdeaudhuij, and C. Vandelanotte, "Are health behavior change interventions that use online social networks effective? a systematic review," *Journal of medical Internet research*, vol. 16, no. 2, p. e40, 2014.

[61] S. A. Moorhead, D. E. Hazlett, L. Harrison, J. K. Carroll, A. Irwin, and C. Hoving, "A new dimension of health care: systematic review of the uses, benefits,

and limitations of social media for health communication," *Journal of medical Internet research*, vol. 15, no. 4, p. e85, 2013.

[62] B. L. Neiger, R. Thackeray, S. A. Van Wagenen, C. L. Hanson, J. H. West, M. D. Barnes, and M. C. Fagen, "Use of social media in health promotion: purposes, key performance indicators, and evaluation metrics," *Health promotion practice*, vol. 13, no. 2, pp. 159–164, 2012.

[63] A. Sarker, R. Ginn, A. Nikfarjam, K. O'Connor, K. Smith, S. Jayaraman, T. Upadhaya, and G. Gonzalez, "Utilizing social media data for pharmacovigilance: a review," *Journal of biomedical informatics*, vol. 54, pp. 202–212, 2015.

[64] R. Sloane, O. Osanlou, D. Lewis, D. Bollegala, S. Maskell, and M. Pirmohamed, "Social media and pharmacovigilance: a review of the opportunities and challenges," *British journal of clinical pharmacology*, vol. 80, no. 4, pp. 910–920, 2015.

[65] E. Velasco, T. Agheneza, K. Denecke, G. Kirchner, and T. Eckmanns, "Social media and internet-based data in global systems for public health surveillance: a systematic review," *The Milbank Quarterly*, vol. 92, no. 1, pp. 7–33, 2014.

[66] T. Webb, J. Joseph, L. Yardley, and S. Michie, "Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy," *Journal of medical Internet research*, vol. 12, no. 1, p. e4, 2010.

[67] L. Sinnenberg, A. M. Buttenheim, K. Padrez, C. Mancheno, L. Ungar, and R. M. Merchant, "Twitter as a tool for health research: a systematic review," *American journal of public health*, vol. 107, no. 1, pp. e1–e8, 2017.

[68] J. B. Colditz, K.-H. Chu, S. L. Emery, C. R. Larkin, A. E. James, J. Welling, and B. A. Primack, "Toward real-time infoveillance of twitter health messages," *American journal of public health*, vol. 108, no. 8, pp. 1009–1014, 2018.

[69] S. Chen, Q. Xu, J. Buchenberger, A. Bagavathi, G. Fair, S. Shaikh, and S. Krishnan, "Dynamics of health agency response and public engagement in public health emergency: A case study of cdc tweeting patterns during the 2016 zika epidemic," *JMIR public health and surveillance*, vol. 4, no. 4, p. e10827, 2018.

[70] V. Hall, W. L. Walker, N. P. Lindsey, J. A. Lehman, J. Kolsin, K. Landry, I. B. Rabe, S. L. Hills, M. Fischer, J. E. Staples, *et al.*, "Update: noncongenital zika virus disease casesâ50 us states and the district of columbia, 2016," *Morbidity and Mortality Weekly Report*, vol. 67, no. 9, p. 265, 2018.

[71] M. Dredze, D. A. Broniatowski, and K. M. Hilyard, "Zika vaccine misconceptions: A social media analysis," *Vaccine*, vol. 34, no. 30, p. 3441, 2016.

[72] A. M. Jamison, D. A. Broniatowski, and S. C. Quinn, “Malicious actors on twitter: A guide for public health researchers,” *American journal of public health*, vol. 109, no. 5, pp. 688–692, 2019.

[73] C. Valentini, “Is using social media “good“ for the public relations profession? a critical reflection,” *Public Relations Review*, vol. 41, no. 2, pp. 170 – 177, 2015. Digital Publics.

[74] T. A. Hadi and K. Fleshler, “Integrating social media monitoring into public health emergency response operations,” *Disaster medicine and public health preparedness*, vol. 10, no. 5, pp. 775–780, 2016.

[75] K.-W. Fu, H. Liang, N. Saroha, Z. T. H. Tse, P. Ip, and I. C.-H. Fung, “How people react to zika virus outbreaks on twitter? a computational content analysis,” *American journal of infection control*, vol. 44, no. 12, pp. 1700–1702, 2016.

[76] E. M. Glowacki, A. J. Lazard, G. B. Wilcox, M. Mackert, and J. M. Bernhardt, “Identifying the public’s concerns and the centers for disease control and prevention’s reactions during a health crisis: an analysis of a zika live twitter chat,” *American journal of infection control*, vol. 44, no. 12, pp. 1709–1711, 2016.

[77] R. Lowe, C. Barcellos, P. Brasil, O. G. Cruz, N. A. Honório, H. Kuper, and M. S. Carvalho, “The zika virus epidemic in brazil: from discovery to future implications,” *International journal of environmental research and public health*, vol. 15, no. 1, p. 96, 2018.

[78] A. R. Daughton and M. J. Paul, “Identifying protective health behaviors on twitter: Observational study of travel advisories and zika virus,” *Journal of medical Internet research*, vol. 21, no. 5, p. e13090, 2019.

[79] S. Masri, J. Jia, C. Li, G. Zhou, M.-C. Lee, G. Yan, and J. Wu, “Use of twitter data to improve zika virus surveillance in the united states during the 2016 epidemic,” *BMC public health*, vol. 19, no. 1, p. 761, 2019.

[80] M. Miller, T. Banerjee, R. Muppalla, W. Romine, and A. Sheth, “What are people tweeting about zika? an exploratory study concerning its symptoms, treatment, transmission, and prevention,” *JMIR public health and surveillance*, vol. 3, no. 2, p. e38, 2017.

[81] B. W. Silverman, *Density estimation for statistics and data analysis*. Routledge, 2018.

[82] P. Du, W. A. Kibbe, and S. M. Lin, “Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching,” *Bioinformatics*, vol. 22, no. 17, pp. 2059–2065, 2006.

[83] D. Newman, J. H. Lau, K. Grieser, and T. Baldwin, "Automatic evaluation of topic coherence," in *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 100–108, Association for Computational Linguistics, 2010.

[84] S. N. Kim, T. Baldwin, and M.-Y. Kan, "Evaluating n-gram based evaluation metrics for automatic keyphrase extraction," in *Proceedings of the 23rd international conference on computational linguistics*, pp. 572–580, Association for Computational Linguistics, 2010.

[85] W. H. O. (WHO), "Who statement on the first meeting of the international health regulations (2005) (ihr 2005) emergency committee on zika virus and observed increase in neurological disorders and neonatal malformations." https://bit.ly/2t7Imzr, 2 2016. Accessed January 2019.

[86] W. H. O. (WHO), "Who statement on the first meeting of the international health regulations (2005) (ihr 2005) emergency committee on zika virus and observed increase in neurological disorders and neonatal malformations." https://bit.ly/2WJjE6b, 2 2016. Accessed January 2019.

[87] A. Attaran *et al.*, "Off the podium: why public health concerns for global spread of zika virus means that rio de janeiro's 2016 olympic games must not proceed," *Harvard Public Health Review*, 2016.

[88] W. H. O. (WHO), "Who statement on the first meeting of the international health regulations (2005) (ihr 2005) emergency committee on zika virus and observed increase in neurological disorders and neonatal malformations." https://www.who.int/en/news-room/detail/28-05-2016-who-public-health-advice-regarding-the-olympics-and-zika-virus, 2 2016. Accessed January 2019.

[89] P. Bennett, K. Calman, S. Curtis, and D. Fischbacher-Smith, *Risk communication and public health.* Oxford University Press, 2010.

[90] B. G. Southwell, S. Dolina, K. Jimenez-Magdaleno, L. B. Squiers, and B. J. Kelly, "Zika virus–related news coverage and online behavior, united states, guatemala, and brazil," *Emerging infectious diseases*, vol. 22, no. 7, p. 1320, 2016.

[91] A. Khatua and A. Khatua, "Immediate and long-term effects of 2016 zika outbreak: A twitter-based study," in *2016 IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom)*, pp. 1–6, IEEE, 2016.

[92] M. J. Paul and M. Dredze, "Social monitoring for public health," *Synthesis Lectures on Information Concepts, Retrieval, and Services*, vol. 9, no. 5, pp. 1–183, 2017.

[93] B. Rezaallah, D. J. Lewis, C. Pierce, H.-F. Zeilhofer, and B.-I. Berg, "Social media surveillance of multiple sclerosis medications used during pregnancy and breastfeeding: Content analysis," *J Med Internet Res*, vol. 21, p. e13003, Aug 2019.

[94] A. Stefanidis, E. Vraga, G. Lamprianidis, J. Radzikowski, P. L. Delamater, K. H. Jacobsen, D. Pfoser, A. Croitoru, and A. Crooks, "Zika in twitter: temporal variations of locations, actors, and concepts," *JMIR public health and surveillance*, vol. 3, no. 2, p. e22, 2017.

[95] M. Farhadloo, K. Winneg, M.-P. S. Chan, K. H. Jamieson, and D. Albarracin, "Associations of topics of discussion on twitter with survey measures of attitudes, knowledge, and behaviors related to zika: probabilistic study in the united states," *JMIR public health and surveillance*, vol. 4, no. 1, p. e16, 2018.

[96] J. P. Guidry, Y. Jin, C. A. Orr, M. Messner, and S. Meganck, "Ebola on instagram and twitter: How health organizations address the health crisis in their social media engagement," *Public Relations Review*, vol. 43, no. 3, pp. 477–486, 2017.

[97] s. chen, Q. Xu, J. Buchenberger, A. Bagavathi, G. Fair, S. Shaikh, and S. Krishnan, "Dynamics of health agency response and public engagement during public health emergency: A case study of cdc tweeting pattern during 2016 zika epidemic," *JMIR Public Health and Surveillance*, vol. 4, p. e10827, 11 2018.

[98] D. Butler, "When google got flu wrong: Us outbreak foxes a leading web-based method for tracking seasonal flu," *Nature*, vol. 494, no. 7436, pp. 155–157, 2013.

[99] L. Simonsen, J. R. Gog, D. Olson, and C. Viboud, "Infectious disease surveillance in the big data era: towards faster and locally relevant systems," *The Journal of infectious diseases*, vol. 214, no. suppl_4, pp. S380–S385, 2016.

[100] L. Zhao, F. Chen, J. Dai, T. Hua, C.-T. Lu, and N. Ramakrishnan, "Unsupervised spatial event detection in targeted domains with applications to civil unrest modeling," *PloS one*, vol. 9, no. 10, p. e110206, 2014.

[101] "Social media fact sheet." https://www.pewresearch.org/internet/fact-sheet/social-media/, note = Accessed: 2020-01-15.

[102] B. Seymour, R. Getman, A. Saraf, L. H. Zhang, and E. Kalenderian, "When advocacy obscures accuracy online: digital pandemics of public health misinformation through an antifluoride case study," *American journal of public health*, vol. 105, no. 3, pp. 517–523, 2015.

[103] D. Dhaliwal and C. A. Mannion, "Anti-vaccine messages on facebook: A preliminary audit," 2019.

[104] A. M. Jamison, D. A. Broniatowski, M. Dredze, Z. Wood-Doughty, D. Khan, and S. C. Quinn, "Vaccine-related advertising in the facebook ad archive," *Vaccine*, vol. 38, no. 3, pp. 512–520, 2020.

[105] D. Albarracin, D. Romer, C. Jones, K. H. Jamieson, and P. Jamieson, "Misleading claims about tobacco products in youtube videos: experimental effects of misinformation on unhealthy attitudes," *Journal of medical Internet research*, vol. 20, no. 6, p. e229, 2018.

[106] W.-y. S. Chou, A. Prestin, C. Lyons, and K.-y. Wen, "Web 2.0 for health promotion: reviewing the current evidence," *American journal of public health*, vol. 103, no. 1, pp. e9–e18, 2013.

[107] M. Lowry and D. Fouse, "Communicating research in an era of misinformation," 2019.

[108] A. Signorini, A. M. Segre, and P. M. Polgreen, "The use of twitter to track levels of disease activity and public concern in the us during the influenza a h1n1 pandemic," *PloS one*, vol. 6, no. 5, p. e19467, 2011.

[109] C. A. Bail, "Emotional feedback and the viral spread of social media messages about autism spectrum disorders," *American journal of public health*, vol. 106, no. 7, pp. 1173–1180, 2016.

[110] L. Safarnejad, Q. Xu, Y. Ge, A. Bagavathi, S. Krishnan, and S. Chen, "Identifying influential factors on discussion dynamics of emerging health issues on social media: A computational study," *JMIR Public Health Surveillance*, 2020.

[111] E. K. Vraga and L. Bode, "I do not believe you: how providing a source corrects health misperceptions across social media platforms," *Information, Communication & Society*, vol. 21, no. 10, pp. 1337–1353, 2018.

[112] A. Gesser-Edelsburg, A. Diamant, R. Hijazi, and G. S. Mesch, "Correcting misinformation by health organizations during measles outbreaks: A controlled experiment," *PloS one*, vol. 13, no. 12, p. e0209505, 2018.

[113] A. Gough, R. F. Hunter, O. Ajao, A. Jurek, G. McKeown, J. Hong, E. Barrett, M. Ferguson, G. McElwee, M. McCarthy, *et al.*, "Tweet for behavior change: using social media for the dissemination of public health messages," *JMIR public health and surveillance*, vol. 3, no. 1, p. e14, 2017.

[114] A. Park, J. Bowling, G. Shaw, C. Li, and S. Chen, "Adopting social media for improving health opportunities and challenges," *North Carolina medical journal*, vol. 80, no. 4, pp. 240–243, 2019.

[115] W.-Y. S. Chou, A. Oh, and W. M. Klein, "Addressing health-related misinformation on social media," *Jama*, vol. 320, no. 23, pp. 2417–2418, 2018.

[116] F. Jin, W. Wang, L. Zhao, E. R. Dougherty, Y. Cao, C.-T. Lu, and N. Ramakrishnan, "Misinformation propagation in the age of twitter.," *IEEE Computer*, vol. 47, no. 12, pp. 90–94, 2014.

[117] S. Jain, V. Sharma, and R. Kaushal, "Towards automated real-time detection of misinformation on twitter," in *2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pp. 2015–2020, IEEE, 2016.

[118] S. Qi, L. AlKulaib, and D. A. Broniatowski, "Detecting and characterizing bot-like behavior on twitter," in *International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation*, pp. 228–232, Springer, 2018.

[119] S. O. Søe, "Algorithmic detection of misinformation and disinformation: Gricean perspectives," *Journal of Documentation*, vol. 74, no. 2, pp. 309–332, 2018.

[120] Y. Wang, F. Ma, Z. Jin, Y. Yuan, G. Xun, K. Jha, L. Su, and J. Gao, "Eann: Event adversarial neural networks for multi-modal fake news detection," in *Proceedings of the 24th acm sigkdd international conference on knowledge discovery & data mining*, pp. 849–857, ACM, 2018.

[121] A. Chadwick and C. Vaccari, "News sharing on uk social media: Misinformation, disinformation, and correction," 2019.

[122] M. Gomez-Rodriguez, J. Leskovec, and A. Krause, "Inferring networks of diffusion and influence," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 5, no. 4, p. 21, 2012.

[123] H. Kwak, C. Lee, H. Park, and S. Moon, "What is twitter, a social network or a news media?," in *Proceedings of the 19th international conference on World wide web*, pp. 591–600, AcM, 2010.

[124] Y. Wang, M. McKee, A. Torbica, and D. Stuckler, "Systematic literature review on the spread of health-related misinformation on social media," *Social Science & Medicine*, p. 112552, 2019.

[125] S. Goel, A. Anderson, J. Hofman, and D. J. Watts, "The structural virality of online diffusion," *Management Science*, vol. 62, no. 1, pp. 180–196, 2015.

[126] M. E. Newman, "Modularity and community structure in networks," *Proceedings of the national academy of sciences*, vol. 103, no. 23, pp. 8577–8582, 2006.

[127] V.-M. Cao-Lormeau, A. Blake, S. Mons, S. Lastère, C. Roche, J. Vanhomwegen, T. Dub, L. Baudouin, A. Teissier, P. Larre, *et al.*, "Guillain-barré syndrome outbreak associated with zika virus infection in french polynesia: a case-control study," *The Lancet*, vol. 387, no. 10027, pp. 1531–1539, 2016.

[128] OSoMe, "botometer." https://botometer.iuni.iu.edu/#!/, 2020.

[129] S. Vijaykumar, G. Nowak, I. Himelboim, and Y. Jin, "Virtual zika transmission after the first us case: who said what and how it spread on twitter," *American journal of infection control*, vol. 46, no. 5, pp. 549–557, 2018.

[130] H. Webb and M. Jirotka, "Nuance, societal dynamics, and responsibility in addressing misinformation in the post-truth era: Commentary on lewandowsky, ecker, and cook," *Journal of Applied Research in Memory and Cognition*, vol. 6, no. 4, 2017.

[131] S. R. Bedrosian, C. E. Young, C. J. D. Smith, Laura A., C. Manning, and e. a. Pechta, Laura, "Lessons of risk communication and health promotion - west africa and united states," *MMWR*, vol. 65, no. 3, pp. 68–74, 2016.

[132] A. Bessi, M. Coletto, G. A. Davidescu, A. Scala, G. Caldarelli, and W. Quattrociocchi, "Science vs conspiracy: Collective narratives in the age of misinformation," *PloS one*, vol. 10, no. 2, p. e0118093, 2015.

[133] G. Eysenbach, "Infodemiology and infoveillance tracking online health information and cyberbehavior for public health," *American Journal of Preventive Medicine*, vol. 40, no. 5, pp. S154–S158, 2011.

[134] A. E. Fard, M. Mohammadi, Y. Chen, and B. Van de Walle, "Computational rumor detection without non-rumor: A one-class classification approach," *IEEE Transactions on Computational Social Systems*, vol. 6, no. 5, pp. 830–846, 2019.

[135] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake news detection on social media: A data mining perspective," *ACM SIGKDD Explorations Newsletter*, vol. 19, no. 1, pp. 22–36, 2017.

[136] V. Qazvinian, E. Rosengren, D. R. Radev, and Q. Mei, "Rumor has it: Identifying misinformation in microblogs," in *Proceedings of the conference on empirical methods in natural language processing*, pp. 1589–1599, Association for Computational Linguistics, 2011.

[137] F. Yang, Y. Liu, X. Yu, and M. Yang, "Automatic detection of rumor on sina weibo," in *Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics*, pp. 1–7, 2012.

[138] S. Hamidian and M. T. Diab, "Rumor detection and classification for twitter data," *arXiv preprint arXiv:1912.08926*, 2019.

[139] A. Friggeri, L. Adamic, D. Eckles, and J. Cheng, "Rumor cascades," in *Eighth International AAAI Conference on Weblogs and Social Media*, 2014.

[140] M. Del Vicario, A. Bessi, F. Zollo, F. Petroni, A. Scala, G. Caldarelli, H. E. Stanley, and W. Quattrociocchi, "The spreading of misinformation online," *Proceedings of the National Academy of Sciences*, vol. 113, no. 3, pp. 554–559, 2016.

[141] A. Tong, D.-Z. Du, and W. Wu, "On misinformation containment in online social networks," in *Advances in neural information processing systems*, pp. 341–351, 2018.

[142] A. Bovet and H. A. Makse, "Influence of fake news in twitter during the 2016 us presidential election," *Nature communications*, vol. 10, no. 1, pp. 1–14, 2019.

[143] L. Safarnejad, Q. Xu, Y. Ge, A. Bagavathi, S. Krishnan, and S. Chen, "Contrasting real and misinformation dissemination network structures on social media during the 2016 zika epidemic," *American Journal of Public Health*, vol. in press, 2020.

[144] Y. R. Tausczik and J. W. Pennebaker, "The psychological meaning of words: Liwc and computerized text analysis methods," *Journal of language and social psychology*, vol. 29, no. 1, pp. 24–54, 2010.

[145] Q. Le and T. Mikolov, "Distributed representations of sentences and documents," in *International conference on machine learning*, pp. 1188–1196, 2014.

[146] "Kaggle." https://www.kaggle.com/, note = Accessed: 2020-06-22.

[147] "Fake news competition on kaggle." https://www.kaggle.com/search?q=fake+news+in%3Acompetitions, note = Accessed: 2020-06-22.

[148] "Fake news prediction - toulouse." https://www.kaggle.com/c/fake-news-toulouse/notebooks, note = Accessed: 2020-06-22.

[149] K. Shu, S. Wang, and H. Liu, "Beyond news contents: The role of social context for fake news detection," in *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, pp. 312–320, 2019.

[150] X. Zhou and R. Zafarani, "Network-based fake news detection: A pattern-driven approach," *ACM SIGKDD Explorations Newsletter*, vol. 21, no. 2, pp. 48–60, 2019.