HAN, YUZHANG, Ph.D. Examining Production, Dissemination, and Consumption of Misinformation: The Case of COVID-19 Pandemic. (2021)
Directed by Dr. Hamid Nemati. 145 pp.

Nowadays, social media is a crucial part of our lives. Platforms like Facebook and Twitter play indispensable roles in the modern information ecosystem, impacting many areas of society. Prevalence of users' speculation and mistrust makes social media a hotbed of *misinformation*, which is information that is wrong or misleading. Misinformation is one of the biggest concerns associated with the use of social media platforms. The COVID-19 pandemic has become a hot topic of misinformation. Huge amounts of misinformation related to the pandemic have been created on social media, covering the public issues such as facial masks, the COVID test and vaccines, and lockdown policies.

One of the consequences of misinformation is opinion polarization, a state in which people are divided into camps such that opinions of people in the same camp are homogenous, while opinions across camps become heterogeneous, even opposite. Social media users with polarized opinions are prone to believing in and spreading misinformation.

The lifecycle of misinformation on social media involves three main components: the *root messages* which contain misinformation, the *producers* who produce the root messages, and the *consumers* who consume the root messages and help spread them further. In this dissertation, I studied these three components' roles in production, dissemination, consumption, and mitigation of misinformation with a focus on the producers of misinformation. Three interrelated research essays have been conducted based on a large, original data set of COVID-19-related misinformation on Twitter. Essay I explores the question: how do producers, root messages, and consumers interact in the production and diffusion of misinformation on social media, and what roles does each of them play? Essay II further anchors on the producers and asks: can producers' communicative intentions, their choice of semantic-linguistic methods, and their polarity of opinion influence the diffusion of misinformation? Finally, Essay III asks: how to reduce misinformation's diffusion by leveraging the knowledge of the producers, consumers, and root messages obtained in Essays I and II using the predictive modeling technology? These essays mainly address the research gap that little research has been focused on the roles of misinformation producers in misinformation diffusion. The research can generate deeper understanding of the mechanism behind misinformation diffusion.

EXAMINING PRODUCTION, DISSEMINATION, AND CONSUMPTION OF MISINFORMATION: THE CASE OF COVID-19 PANDEMIC

by

Yuzhang Han

A Dissertation

Submitted to

the Faculty of The Graduate School at

The University of North Carolina at Greensboro

in Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

Greensboro

2021

Approved by

Dr. Hamid Nemati

Committee Chair

APPROVAL PAGE

This dissertation written by Yuzhang Han has been approved by the following committee of the Faculty of The Graduate School at The University of North Carolina at Greensboro.

Committee Chair	Dr. Hamid Nemati
Committee Members	Dr. Indika Dissanayake
	Dr. Jiyong Park
	Dr. Minoo Modaresnezhad
	Dr. Nikhil Mehta

6/24/2021
Date of Acceptance by Committee
6/14/2021
Date of Final Oral Examination

ACKNOWLEDGEMENTS

My sincere gratitude to Dr. Hamid Nemati, my advisor and committee chair. I truly appreciate all his guidance and motivation on my dissertation and my pursuit of the doctoral degree.

I also would like to appreciate my dissertation committee, Dr. Indika Dissanayake, Dr. Jiyong Park, Dr. Minoo Modaresnezhad, and Dr. Nikhil Mehta, for their instructions and input to my dissertation and doctoral research work.

Many thanks to everyone—professors, staff, peers…—who gave me such lots of help on my dissertation and doctoral study.

Special thanks to Alta Zhang, who has offered tremendous support for the completion of this work.

PREFACE

Nowadays, social media is a crucial part of our lives. Platforms like Facebook and Twitter play indispensable roles in the modern information ecosystem, impacting many areas of society. Social media creates a disintermediated and decentralized environment for the production and dissemination of contents. The ease by which content is created and disseminated can encourage speculation and may lead to mistrust among the users. Prevalence of users' speculation and mistrust makes social media a hotbed of *misinformation*, which is information that is wrong or misleading. Misinformation is one of the biggest concerns associated with the use of social media platforms. As a result, it has garnered much interest from researchers and practitioners alike. Social media contents with misinformation that includes fabricated ideas, incomplete depiction of events, or misleading interpretation of facts can distort people's understanding of the events, creating misguided beliefs among people. Contents that include misinformation tend to spread broader and faster than authentic contents. Once misinformation spreads, it is both difficult and expensive to correct or contain its damage. Many issues of public interest have been affected by misinformation, such as digital currencies, natural disasters, global warming, and vaccination. In particular, the COVID-19 pandemic has become a hot topic of misinformation. During this crisis, huge attention of the public has been attracted. Any news event, such as the stay-home order and development of vaccines, can stir extensive debate on social media. In turn, huge amounts of misinformation have been created on social media in many different forms, such as fake news, conspiracy theories, and hoax.

One of the consequences of misinformation is *opinion polarization*, a state in which people are divided into camps such that opinions of people in the same camp are homogenous, while opinions across camps become heterogeneous, even opposite. Opinion polarization can impact many aspects of our lives. It can distort the social relationships among us (such as our friendships and marriages) and affect our economic, social, and political behaviors. Users with polarized opinions are not only prone to believing in misinformation. They are also more likely to spread it, which renders opinion polarization both a cause and a consequence of misinformation diffusion.

The lifecycle of misinformation on social media involves three main components: the *root messages* which contain misinformation, the *producers* who produce the root messages, and

the *consumers* who consume the root messages and help spread them further. To reduce the diffusion of misinformation, it is crucial to understand the roles the producers play in not only framing the root messages but also the behavioral changes of consumers. Damage of misinformation can be mitigated by finding out which root messages contain misinformation predictively. Completion of this mitigating task relies on understanding the associations between consumers and producers. Their associations need to be examined from producers' as well as consumers' perspectives. These two divergent perspectives include consumer's characteristics, such as their account demographics, social relations, and past activities and producers' characteristics including the origins of their opinions, communicative intentions, and the semantic-linguistic methods they use during production of misinformation.

In this dissertation, I studied these three components' roles in production, dissemination, consumption, and mitigation of misinformation with a focus on the producers of misinformation. Three interrelated research essays have been conducted based on a large, original data set of COVID-19-related misinformation on Twitter. Essay I explores the question: how do producers, root messages, and consumers interact in the production and diffusion of misinformation on social media, and what roles does each of them play? Essay II further anchors on the producers and asks: can producers' communicative intentions, their choice of semantic-linguistic methods, and their polarity of opinion influence the diffusion of misinformation? Finally, Essay III asks: how to reduce misinformation's diffusion by leveraging the knowledge about the producers, consumers, and root messages obtained in Essays I and II using the predictive modeling technology? These essays mainly address the research gap that little research has been focused on the roles of misinformation producers in misinformation diffusion. The research can generate deeper understanding of the mechanism behind misinformation diffusion. This understanding can help us create more powerful solutions for warning the platforms and users of their risk to be harmed by misinformation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I: INTRODUCTION

Social media have been widely recognized for its capability to democratize online conversation (Ferrara, 2017). Platforms like Twitter and Facebook play an indispensable role in the modern information ecosystem, impacting many areas of society such as political procedures (Allcott & Gentzkow, 2017; Anstead & O'Loughlin, 2015), civil movements (de Waal & Ibreck, 2013; Frangonikolopoulos & Chapsos, 2012), public health interventions (Kass-Hout & Alhinnawi, 2013; Thackeray et al., 2012), and education (Gikas & Grant, 2013; Selwyn, 2012). However, this powerful mechanism can also be abused for malicious purposes: extremist groups use social media for propagating violence online and recruiting new cadres (Awan, 2017); stock market manipulators have concerted operations to create upheavals on financial systems (Ferrara, 2015a). Social media is also the hotbed of various harmful, or even criminal, activities such as identity theft, cyberstalking/bullying scams, and purchasing illegal items (Irshad & Soomro, 2018).

In this dissertation, I examine one of the greatest concerns about social media that has been raised for democratic societies—the rampage of misinformation on social media platforms (Ferrara, 2015b; Marwick & Lewis, 2017; Tucker et al., 2018). Misinformation refers to any wrong information or misleading information (Stahl, 2006). It became a hot topic for social media research when social media was a still fresh term in the media vocabulary (Chamberlain, 2010; Oh et al., 2010). During the 2016 U.S. presidential election, misinformation was brought to the public attention in a new, intentional form—fake news. Its devastating impact on the public perception of truth and false and people's fundamental belief in media not only stirred the atmosphere of that election, but has reached to many issues we care about today: Misinformation on social media has been reported to target participants of the Black Lives Matter movement (Marwick & Lewis, 2017). Misinformation-based conspiracies are found deeply tangled with both pro- and anti-vaccination discussions (Chiou & Tucker, 2018). Terrifying but fabricated events, such as outbreaks of Ebola in Atlanta, an explosion of chemical plan in Louisiana, and nuclear plant accidents in Ukraine, were spread virally on social media in the form of misinformation (Linvill & Warren, 2020).

The COVID-19 pandemic, being one of the most severe public health challenges in human history and has become an arena of all sorts of misinformation on social media. During

this crisis, people's activities responding to the pandemic both on the ground and online are becoming more and more simultaneous and intertwined. Social media provides an unprecedented opportunity for the public to share and produce authentic and false information that might impact the health of everyone (Abd-Alrazaq et al., 2020). To counter this pandemic of misinformation, main platforms such as Twitter and Facebook invited the professional fact checkers of social media content to enhance the information validation functionalities of the platforms (BBC News, 2020; Guy Rosen, 2020; Sophie Lewis, 2020). Nevertheless, significant, fast-growing volumes of COVID-related messages of misinformation still have been captured on various social media platforms (Brennen et al., 2020). On some newly emerging platforms like Gab, the volume of unreliable messages is even up to 70% of reliable messages, while the amount of engagements for unreliable messages is about 300% larger than the amount for reliable ones (Cinelli et al., 2020). As a byproduct, the number of English-language fact-checkers increased more than 900% from January 2020 to March 2020, which implies that the total size of all kinds of misinformation related to coronavirus had almost certainly risen even faster (Brennen et al., 2020).

There are quite a few hot topics of misinformation which have been wildly propagated during the pandemic, such as Bill Gates' plan to microchip individuals and force them into vaccination (Georgiou et al., 2020), and 5G facilities being related to the spread of coronavirus (Ahmed et al., 2020). Many social media users have been deceived by such information. Take the 5G fiction as an example: among the sampled Twitter users who used the hashtag #5GCoronavirus, 34.8% of them expressed views that 5G and COVID-19 were linked (Ahmed et al., 2020). People's stress under the pandemic and their pre-existing conspiracy beliefs together contribute to their beliefs in COVID-19-related misinformation (Georgiou et al., 2020).

There are two types of definitions of misinformation. Under the first type of definition, misinformation is any information that is incorrect (Karlova & Fisher, 2013), and as the Oxford English Dictionary suggests, wrong information or misleading information (Stahl, 2006). In the social media context, misinformation is an umbrella term that includes all false or inaccurate information spreading on social media platforms (Wu et al., 2019); it is any claim of fact that is not true due to lack of scientific support (Chou et al., 2018; Kouzy et al., 2020). By contrast, the second type of definition describes misinformation as unintentionally false information, as compared to *disinformation*, which is intentionally false (Jack, 2017; Quandt et al., 2019; Torres

et al., 2018); misinformation is made without any intention to mislead (Lewandowsky et al., 2012). In this study, considering that the first type of definition has been applied more widely, I follow that type of definition and use the term misinformation to refer to any false information on social media no matter if it was created to mislead the audience purposefully.

Diffusion of misinformation involves three components: (1) the initial messages that contain misinformation, denoted as *root messages* in this study (Bharadwaj & Shao, 2019; Shu, Sliva, et al., 2017); (2) *producers* of root messages, who create the messages on social media by composing them and posting them using their accounts (Parikh & Atrey, 2018; Wang et al., 2019); (3) *consumers* of root messages who receive and then spread the messages to other accounts by, for example, retweeting/sharing, replying to/commenting on, or quoting them (Vraga et al., 2020; M. C. Wagner & Boczkowski, 2019). In addition, users who receive but do not react to the root messages are called *receivers* (in other words, non-consumers) in this study. If reviewed via the lens of these components, the research literature on misinformation on social media has been mainly focused on the component of root messages. In particular, efforts have been made to explore the methods for detecting misinformation, especially fake news. One type of detection methods mainly rely on the message content itself (Shu, Sliva, et al., 2017), extracting various linguistic clues from the message bodies and titles (Aldwairi & Alwahedi, 2018; Pérez-Rosas et al., 2017). Another, more advanced type of detection methods focus on the social content of misinformation, leveraging the propagation patterns of the root messages (Yang Liu & Wu, 2018; Wu & Liu, 2018). Machine learning techniques are used to convert knowledge about the messages and their social content into prediction of message veracity (Shu et al., 2018).

Another component, the consumers of misinformation, has attracted a relatively smaller amount of interest. Studies in this area mainly draw on two factors—consumers' characteristics and the way consumers handle misinformation—in order to find out how these factors are associated with users' consumption behavior of misinformation. For the first factor, it has been found that user characteristics, such as having long-standing accounts, making fewer posts, and expressing more favor actions to others' posts (e.g., giving more "likes"), are associated with higher risk of trusting and spreading misinformation (Shu et al., 2018). For the second factor, it was found that, users' acceptance of misinformation differs based on their trust in news sources and their personal involvement in the information on social media (Flintham et al., 2018).

Compared to the aforementioned two components, study of producers' characteristics has attracted disproportionally lower attention. A part of research in this area deals with the nature of producers and have identified different types of producers of misinformation, such as internet trolls, conspiracy theorists, and hyper-partisan news outlets; these producers differ in their motivations, techniques used, target platforms, and production outcome (Marwick & Lewis, 2017). In particular, social bots were found to play a key role in the spread of misinformation, such as fake news (Shao et al., 2017). Based on the understanding of the nature of misinformation producers, some research has been done to detect accounts that create misinformation and other types of malicious information using statistical analysis and predictive modeling (Boshmaf, Logothetis, et al. 2015; Ercsahin et al. 2017). Finally, a small amount of research is focused on the approaches taken by the producers to producing misinformation. For instance, they can debate actively with genuine users, or be engaged in hate speech or other forms of online harassment. They can spread a variety of contents, such as fake videos, blogs, memes, or pictures (Bradshaw & Howard, 2018; Marwick & Lewis, 2017).

A severe social-political issue associated with both the consumers and producers of misinformation is *opinion polarization*. It refers to a status that users in different communities within a social network present distinct, even opposite opinions on a common topic; different communities tend to hold negative attitudes towards each other (Conover et al., 2011; DiMaggio et al., 1996; Maes & Bischofberger, 2015). Polarization on social media has been identified as a pressing social problem ever since the main social media platforms such as Twitter and Facebook gained mass popularity at the beginning of 2010s. Polarization has been found to not only disturb democratic political systems, but also affect the social relations we seek to enter into, such as our friendships, romantic relationships, or marriages (Iyengar et al., 2019; Nicholson et al., 2016). Furthermore, more and more evidence has been found which renders polarization of social networks as a predictor of the prevalence of misinformation (Allcott & Gentzkow, 2017; Bessi et al., 2016; Tucker et al., 2018; Vicario et al., 2019). On the one hand, users in a polarized network are more prone to diffuse false information (Bessi et al., 2016). On the other hand, polarized users tend to believe misinformation because their vulnerability to false information is increased by directionally motivated reasoning, which occurs when their belief is biased due to polarization (Tucker et al., 2018).

Practically, knowledge about the components of misinformation diffusion can be used for mitigating the damage of misinformation by leveraging advanced technologies such as predictive modeling (Oshikawa et al., 2018). Almost all the efforts on this track have been devoted to the root messages component, trying to detect misinformation messages out of true information (Bondielli & Marcelloni, 2019; X. Zhou & Zafarani, 2018). In comparison, only little research takes the consumers or producers as the detection targets, namely, predicting the potential producers or warning users at risk of being harmed by misinformation. Furthermore, the existing research targeting consumers is limited to only considering characteristics of the users who receive misinformation messages, without making good use of the characteristics of root messages and producers (Boshmaf, Logothetis, et al., 2015; Boshmaf, Ripeanu, et al., 2015; Pennycook & Rand, 2020; Shen et al., 2019; Shu, Zhou, et al., 2019; Tjostheim & Waterworth, 2020).

Observations above reveal a series of gaps in the existing research: First of all, production and diffusion of misinformation on social media is driven by a complicated mechanism. There could exist countless associations among the characteristics and behavior of the misinformation producers, root messages, and consumers, which influence the nature and propagation of the root messages produced. However, only limited efforts have been made to reveal the basic mechanism that drives the diffusion of misinformation. For example, what characteristics of a producer determine that his or her root messages can be propagated virally? What characteristics of a consumer determine that the consumer would consume this piece of false message, rather than that piece of message? Although work has been done to explore each of the single components of misinformation, there still misses an integrative examination of how these components interrelate. Specifically, how root messages are produced, passed by, and consumed, especially with consideration of all the three components and their interactions simultaneously and integrally.

Second, compared to the consumers and root messages, there is rather limited attention paid to the producer side, namely the upstream, of misinformation diffusion. Third, little efforts have been made to scrutinize these producers as humans engaged in *language production*, inspecting three important dimensions of their characteristics during misinformation production from a language production perspective: (1) *ideological dimension* (studied in, e.g., Au et al., 2021; Jost et al., 2018; Spohr, 2017): what opinions they hold upon the issues which the

misinformation messages are about, (2) *pragmatic dimension* (studied in, e.g., Arielli, 2018; Seifert, 2002; Søe, 2017): what communicative intentions they want to satisfy through the messages, or, what they intend to express, and (3) *semantic-syntactic dimension* (studied in, e.g., Granik & Mesyura, 2017; Shu, Sliva, et al., 2017): what semantic-syntactic approaches they apply (intentionally or unintentionally) in the messages, as a result, misleading the viewers. Fourth, prior work has suggested inspecting the mechanism of misinformation through the ideological (Calvillo et al., 2020; Sikder et al., 2020), pragmatic (Parikh & Atrey, 2018; Seifert, 2002), and semantic-syntactic characteristics (Bharadwaj & Shao, 2019; Choudhary & Arora, 2021) of the involved parties. However, given the important role of the misinformation producers, it has not been further investigated how the ideological, pragmatic, and semantic-syntactic characteristics of the producers would impact the other components of misinformation diffusion, namely, the messages themselves and users' reactions to the messages.

Fifth, there is a large amount of research done to detect misinformation from authentic information. However, most of the existing studies on this track mainly rely on information of the root messages and receivers to support the detection (Pérez-Rosas et al., 2017), or even treat producers and receivers as the same group of users who propagate misinformation (Shu, Sliva, et al., 2017). Fewer studies have utilized knowledge about the producers to facilitate the detection. Furthermore, most of these studies focus on features of producers immediately accessible in their online profiles (e.g., Ruchansky et al., 2017; Shu, Wang, et al., 2017, 2019). Producers' intentions to produce the misinformation have not been utilized in misinformation detection.

In this study, I take a perspective of language production to approach the research gaps above. Misinformation production on social media can be seen as an instance of language production (e.g., Hou et al. 2019; Stine and Agarwal 2019). Language production, as presented in Figure 1, is an iterative process consisting of four stages (Willem J. M., 1989): conceptualization, formulation, articulation, and self-monitoring. In the conceptualization stage, the producer (of misinformation or language) conceives of a communicative intention (note that this is the intention of communication, not the intention to deceive), selects the relevant information to be expressed for the realization of this intention, and orders and refines this information for expression. In this phase, the producer is expected to possess declarative knowledge, which refers to the producer's opinion and perception of the world—how she thinks about the issue she is going to address, how she feels about the current environment, etc.

Declarative knowledge constitutes the ideological characteristics of the producers mentioned in the second research gap.

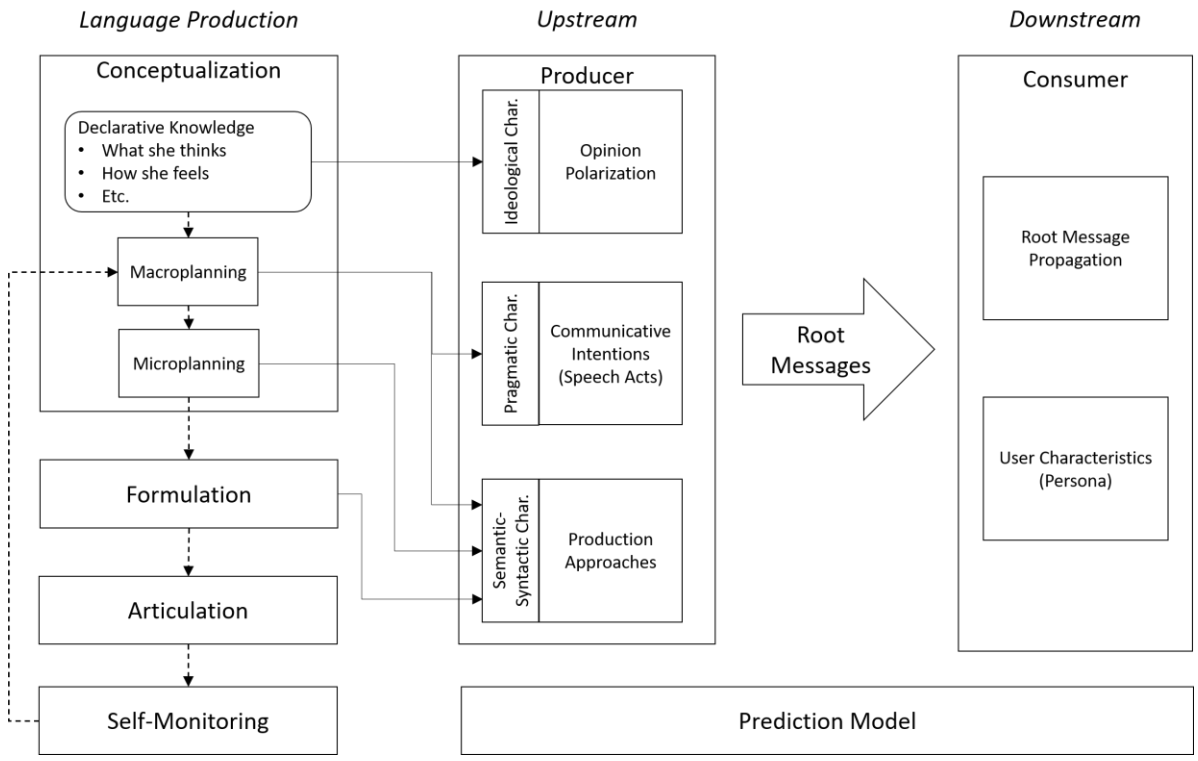

Figure 1. Conceptional Model of Diffusion of Misinformation

In the remaining part of the conceptualization stage, producer's declarative knowledge goes through two procedures (Willem J. M., 1989): macroplanning and microplanning. In the macroplanning procedure, the producer finds out her communicative intention and assembles appropriate information whose expression will reveal the intention to the receiver. Then, in the microplanning procedure, the producer fine-tunes this information by assigning the right propositional shape to each chunk of information, as well as defining the particular topic and focus that will guide the receiver's allocation of attention. As a result of macroplanning, the producer determines the *speech acts*, which are linguistic acts performed in the process of speaking as defined in pragmatics (Ilyas & Khushi, 2012; Sadock, 2004). Speech acts are the functions the producer wants to realize through the current communication, such as suggesting, requesting, promising, and thanking (Ilyas & Khushi, 2012). They serve to convey the communicative intention in a discourse (Searle & Searle, 1969; Villarroel Ordenes et al., 2017;

R. Zhang et al., 2011). Performing a speech act is essentially expressing a certain intention in uttering certain words. In relation to the second research gap, communicative intentions and the selected speech acts reflect the pragmatic characteristics of a producer. Also, misinformation producers can perform particular semantic approaches during macroplanning and microplanning to producing misinformation. For instance, they can completely fabricate units of information, or hide certain parts of the true information (L. Zhou et al., 2004).

Next, in the formulation stage of language production, the producer transforms the selected information from stage one into words and sentences organized by certain syntax, which are ready to be expressed (Willem J. M., 1989). In this stage, misinformation producers can perform particular syntactic approaches to producing misinformation. For instance, they can create sentence vagueness by applying contradictory or impenetrable sentence structures, or increase the message indirectness by making questions followed by questions (L. Zhou et al., 2004). In relation to the second research gap, this stage together with the formulation stage create the semantic-syntactic characteristics of misinformation producers.

In the next stage, articulation, the producer expresses the output from stage three by, for example, speaking it or writing it. In the final stage, self-monitoring, the producer comprehends what was just expressed and goes back to stage one to start producing the next message.

Motivated by all the considerations above, I propose to use this dissertation to mainly examine the upstream side of misinformation diffusion, namely, the producers of misinformation root messages, as the driver of misinformation diffusion. Taking the perspective of language production, the research sheds light on the three dimensions (i.e., ideological, pragmatic, and semantic-syntactic dimension) of producer characteristics mentioned above, inspecting their opinions (represented by polarity or polarization of their opinions), what are their communicative intentions conveyed through root messages (represented by their use of speech acts), and how they express these messages in the form of misinformation (represented by the semantic-syntactic approaches they take to producing misinformation), as well as how these factors might impact people's reaction to the produced root messages. These broad questions will be answered by conducting three interrelated research essays based on Twitter data related to COVID-19. The main research question addressed by each essay is listed below:

- Essay I: How do the three components of misinformation influence each other in the diffusion of misinformation? What roles do they play?

- Essay II: How do producers' communicative intentions, their choice of misinformation production approaches, and their opinion polarization influence the diffusion of misinformation?
- Essay III: How to identify the root messages with misinformation by leveraging knowledge about the producers and other components of misinformation diffusion?

Essay I addresses the first research gap by exploring the *online profiles* of the three components. Essay II addresses the second, third and fourth research gaps. One of the central tasks of this essay is to investigate the roles of producers' communicative intentions, their choice of approaches to producing misinformation, and the polarity of their opinion in the entire process of misinformation diffusion. Two variables of misinformation diffusion are analyzed in order to see how they are influenced by producer intentions and their production approaches. These variables are the way in which the messages propagate and the consumers who spread the messages.

Finally, Essay III aims at the last research gap. In this essay, a framework that detects misinformation is proposed, implemented, and evaluated. The framework is presented following the style in similar works (e.g., Chau et al. 2020; He et al. 2018; Liebman et al. 2019; Zhang and Ram 2020) and is novel in that it leverages the knowledge about misinformation producers obtained in the previous essay, including polarity of producers' opinions, their communicative intentions, and their choice of production approaches.

Structurally, this work is organized as follows: Chapters II, III, and IV present Essays I, II, and III, respectively. For each essay, a literature review, the research model and hypotheses, methodology, and analytic or evaluation results are provided independently. Finally, Chapter V concludes the dissertation.

CHAPTER II: ESSAY I: UNDERSTANDING PRODUCERS, ROOT MESSAGES, AND CONSUMERS OF MISINFORMATION

2.1 Introduction

Essay I focuses on the first gap. This essay establishes the model of misinformation diffusion (illustrated in Figure 1). In the upstream lies the *producer* of misinformation. A producer is fully characterized by his or her online profile—a structured description of who the producer is, how he or she behaves, and who are connected to this user. The producer produces a piece of *root message* that contains misinformation. As an instance of online content, all the properties of the root message are also included in its profile.

In the downstream, the root message is received by the *receiver*, who is typically a follower of the producer's account, and thus is also illustrated by a profile. If the receiver is interested in the root message, either positively or negatively, she will react to the message by, for example, forwarding it, replying to it, or commenting on it. In this case, the receiver becomes a *consumer* of the message. Followers of a consumer will also receive the message. In this way, the message is spread to more and more users in the network.

Accordingly, this essay is mainly engaged with three constructs: the *Producer Profile, Consumer Profile, and Root Message Profile*. Characteristics of producers and consumers are represented by multiple types of features of the social media users who produce or react (e.g., retweeting or replying) to the root messages they received. These features are denoted collectively as Producer or Consumer Profile. Various types of Producer or Consumer Profile are investigated, such as their activeness persona, referring to features related to consumers' activeness on Twitter, and emotion/sentiment persona, referring to features describing consumers' emotion and sentiment expressed in their text. Similarly, social platforms also produce and expose detailed, structured description of instances of online content. In this study, the *Root Message Profile* refer to such description of the root messages. Many attributes are available in Root Messages Profile, such as the content and metadata of the messages, and how the messages were propagated.

Research questions in Essay I are as follows:

Q1: How do the characteristics of misinformation producers reflected in their profiles influence the content and propagation of the root messages they produce?

Q2: How do the characteristics of misinformation producers reflected in their profiles influence which users spread the root messages produced by these producers?
Q3: How do the linguistic characteristics of the text body of root messages influence the propagation of these root messages?
Q4: How do the characteristics of misinformation root messages reflected in these root messages' online profiles influence which users spread these root messages?

2.2 Theoretical Background and Literature Review

Overall, Essay I examines the role of producers in the misinformation diffusion by focusing on their communicative intentions and their choice of production approaches. The essay examines how these two constructs influence the way in which roots messages propagate, and who will consume these messages. In this subsection, the mechanism of misinformation diffusion is delineated. Then, the main constructs involved in each of the relations above are discussed with literature review. Finally, the research model and hypotheses that knit these relations together are proposed.

2.1.1 Diffusion of Misinformation on Social Media and Its Components

2.1.1.1 Diffusion of Misinformation on Social Media

Misinformation on social media appeared as a research topic when social media was just added to the media vocabulary. At that time, research was focused on fundamental issues such as the structures of social networks on specific social media platforms and how these structures might facilitate passing of misinformation accidentally and deliberately (Chamberlain, 2010). It was observed that platforms like Twitter are especially suitable for misinformation operations due to the casual style of communication and asymmetrical structure of social networks.

On this basis, a seminal model of misinformation diffusion was constructed (Karlova & Fisher, 2013), which identifies the two basic roles in misinformation operations—the producers (called diffusers in that work) and receivers. While these two roles have different behavioral patterns, the behavior of producers is guided by their intents that are personality- or socialization-motivated. Connecting the producers and receivers are the root messages of misinformation. The key features (called cues) of the message content and producers can indicate the possibility of deception.

Misinformation on social media became a sensational topic when a species of misinformation—fake news—attracted huge attention during the 2016 U.S. presidential election

(Martens et al., 2018). Fake news refers to deliberately fabricated information presented in the form of news articles (Shu, Sliva, et al., 2017). While other types of misinformation can be casual in presentation, fake news is usually formulated formally and pseudo-professionally. Given the immense speed of spread of fake news and its infectious effects on the receiving population that was shown during the 2016 election, it was suggested that fake news diffusion bears many similarities to epidemic development and can be studied as such (Kucharski, 2016).

Diffusion of misinformation depends on three components: producers of misinformation, root messages of misinformation, and consumers of misinformation (see Figure 1). Producers are the users who create the misinformation and initialize its spreading by posting it on social media. Misinformation is carried by the root messages produced by the producers in the forms of tweets, Facebook pages, Instagram posts etc. Users who receive the root messages can react to them by forwarding them, replying to or commenting on them, or quoting them in their own messages—in these cases they become consumers of the root messages. Typically, these activities enable the some or all of the followers of a consumer to receive the message. In this way, misinformation is spread to more and more users in the same social network, forming a diffusion on social media.

2.1.1.2 Producers

Literature identifies several types of producers of misinformation. Depending on the intentions they have when they produce misinformation, producers could be internet trolls, gamergaters, hate groups and ideologues, conspiracy theorists, hyper-partisan news outlets, and politicians. These producers differ in their motivations, techniques used, target platforms, and production outcome (Marwick & Lewis, 2017). In particular, social bots were found to play a key role in the spread of misinformation. Accounts that actively spread fake news are much more likely to be bots, which are especially active in the early phases of news spreading (Shao et al., 2017).

Activities of producers are also influenced by the platforms they operate on. Producers on mainstream media and social media create different styles of misinformation, and accuse the other group as the cause of misinformation spreading (Al-Rawi, 2019). Also, the same producer can behave differently on different types of platforms. For instance, some producers might be trial ballooning a wide range of messages with different styles on a platform featuring smaller, more engaged communities (such as Reddit), and distribute only the most effective messages to another platform featuring much larger scale but casual users (such as Twitter) (Lukito, 2020).

In addition, different producers can employ distinct approaches to producing misinformation. They can pretend to be commentators on social media who actively debate with genuine users, or act as trolls who are engaged in hate speech or other forms of online harassment. They can produce a variety of contents, such as fake videos, blogs, memes, or pictures. They can also perform malevolent takedown of legitimate content or accounts (Bradshaw & Howard, 2018).

2.1.1.3 Consumers

Consumers are those who receive misinformation messages and react to them (e.g., by retweeting, sharing, replying to, commenting on, or quoting them) in a way that causes the messages to be received by other users. Consumers of misinformation and their consumption activities have drawn substantial research interest. Overall, popular misinformation messages, especially fake news stories, are found to be shared more widely than the most popular mainstream news stories on social media platforms such as Facebook (Allcott & Gentzkow, 2017). Although laypeople on average are good at distinguishing between lower- and higher-quality information sources (Pennycook & Rand, 2019), many users were still reported to believe some kinds of fake news stories. This study focuses on consumers on Twitter, who propagate the root messages they receive by retweeting them. Therefore, they are denoted as *consumers*.

Veracity is defined as the property that a statement truthfully reflects the aspect it is made about (Gollmann, 2012). In misinformation studies, veracity of a message typically refers to if the message is misinformation or authentic (Shu, Sliva, et al., 2017). Users' capability to discern the veracity of messages depends on two factors. The first factor is user characteristics. Several types of user characteristics have been found related to higher possibility to believe misinformation, including having accounts with higher age, making fewer posts, expressing more "favor" actions to others' posts (e.g., giving "likes"), having fewer followers and more followees, and being less extrovert and friendly with other users (Shu et al., 2018). The second factor is the way in which users handle information they receive. It is found that some users tend to consider everything on credible sources (such as mainstream news outlets) be true information, while some users do not care about source credibility when judging on veracity of received information. Also, some users do not examine the veracity of information carefully if they have no personal or professional interest in the information (Flintham et al., 2018). In addition, spread of misinformation can be best explained by users' conformity to others online.

People have an intuition to conform to what other people do (Sunstein, 2002). This intuitive conformity can trigger the echo chamber effect that strengthens the chance of misinformation diffusion among its receivers (Colliander, 2019).

2.1.1.4 Root Messages

Root messages are the output of misinformation production. The root messages this study focuses on are *tweets*, namely, the messages created by the producers. Being the carrier of misinformation, root messages are characterized by several types of features, such as their lifecycles (e.g., creation time, number of likes received), linguistic features (e.g., number of words and sentences, unique words, readability) (Perikos & Hatzilygeroudis, 2016; Sharif et al., 2016; Tan et al., 2014), veracity, emotion and sentiment (e.g., positive, negative, fear, happy, anger, surprise) (Ajao et al., 2019; Newman et al., 2003; Shu, Zhou, et al., 2019), and topics (e.g., masks, stay home order, vaccination) (Bovet & Makse, 2019; Kwon et al., 2013; Vosoughi et al., 2018). These features not only can reflect the psychological and social status of the producers (Cuerie, 1952; Danescu-Niculescu-Mizil et al., 2011; Pennebaker et al., 2015; Soames, 1984), but can also influence how receivers will act on them (Baldwin et al., 2013; Bandari et al., 2012). Also, different root messages can display unique propagation patterns within social networks (Vosoughi et al., 2018; Z. Zhao et al., 2020).

Several aspects of misinformation diffusion can be characterized predictively using predictive modeling. Research has been devoted to detecting misinformation root messages (misinformation detection, Aldwairi and Alwahedi 2018; Ozbay and Alatas 2020; Pérez-Rosas et al. 2017; Shu, Sliva, et al. 2017), producers (suspicious account detection, Er\csahin et al. 2017; Jia et al. 2017), and consumers (victims detection, Boshmaf, Ripeanu, et al. 2015; Guerra et al. 2013; Shen et al. 2019; Wagner et al. 2012).

Note that the terms producers, root messages, and consumers do not only apply for misinformation in this dissertation. When non-misinformation is discussed, producers, root messages, and consumers refer to the producers, root messages, and consumers of non-misinformation.

2.1.2 Producer Profile and Consumer Profile

Producer Profile refers to the personal features that characterize the producers of root messages as humans who use a social media platform. These features describe who these social media users are and how they are engaged in the platform. Prior research suggests that such

personal characteristics of social media users are crucial in helping people understand the diffusion of misinformation. For example, Shu et al. (2019), in their comprehensive survey study, considered user-related features as being able to provide useful information for fake news detection. The user-based features are categorized into individual level features, such as demographics of individual users (account age, follower count, followee (called friends on Twittter) count, number of posts, etc.), and group level features, which are aggregates of individual features per user community. In a more recent study (Vicario et al., 2019), user-related features such as the numbers of comments, likes, and posts per user and per group were employed to detect potential topics of misinformation. In this research, I employ an even broader set of features to represent the Consumer Profile, including features related to consumers' activeness on Twitter, and their sentiment and emotion expressed towards root messages.

This essay focuses on four types of information in producers' profiles on Twitter: (1) Information about their activeness on the platform, such as how many messages they have posted, how many likes they have received, how many accounts they have listed, etc.; (2) producers' socialization on the platform, such as how many friends and followers they have, etc.; (3) text statistics of producers' user description: every user has a piece of self-introduction posted on the account, this is the user description of the user. Statistics of user description include basic facts of this text instance, such as the number of words and its readability score; and (4) sentiment and emotion expressed in producers' user description.

Consumer Profile refers to the same types of information related to the consumers. This study focuses on consumers as the consumers. Therefore, this construct is also denoted as Consumer Profile in this study.

2.1.3 Root Message Profile

Root Message Profile refers to all the information of a root message that is available to, collected by, and partially exposed by the social media platform. This essay focuses on tweets on the Twitter platform. Therefore, Root Message Profile is also denoted as Root Message Profile. There are three types of information in Root Message Profile within the scope of this study: (1) text statistics of the text body of a message (or tweet): basic information about the text body, such as number of words, sentence length, the readability score. (2) sentiment and emotion expressed in the text body. (3) propagation of the root message: how often the message has been passed by, liked, commented on, etc.

Propagation of root messages could be influenced by producer characteristics. It refers to the propagation patterns of root messages. An emerging stream in the research of user consumption activities is the analysis of propagation patterns of misinformation. Propagation of a root message offers topological and dynamic views of users' consumption of the message (Yang Liu & Wu, 2018; Wu & Liu, 2018). Propagation of messages can be visualized as propagation networks, which can be characterized by two types of measures—the scale measures and topological measures. Scale measures mainly refer to the depth, width, and breadth of a propagation network. Study using these measures (Vosoughi et al., 2018) shows that misinformation, especially fake news, propagates significantly farther, faster, deeper, and more broadly than true information regardless the topics, and the effects were more pronounced for political news than for other types of news such as that about terrorism, natural disasters, science, urban legends, or financial information. Alternatively, topological measures include metrics such as the layer ratio (ratio between numbers of re-posters on two layers in the network) and hop-distance between any two consumers in the network (Z. Zhao et al., 2020). Analysis of this type of measures indicates that misinformation spreads with distinctive network topology compared to true information even at early stages. Later adopters, instead of direct followers of the producers, mainly foster the penetration of misinformation in social networks (Z. Zhao et al., 2020).

Propagation patterns of root messages can be understood by analyzing the propagation *cascades* of the messages. A propagation cascade, or cascade, is a star-shaped, unbroken chain of misinformation with a common source that is created when information is shared on social media (Vosoughi et al., 2018; X. Zhou & Zafarani, 2018). Typologically, a cascade is a directed tree representing the propagation of a root message among the producer and consumers. The root of a cascade represents the producer (or the root message), and each node represents a consumer (or her post made to the root message). An edge from nodes n_1 to n_2 means n_2 responds (in this study, "responds" refers to "retweets") to the post by n_1. Figure 2 depicts the topology of cascade for a misinformation root message.

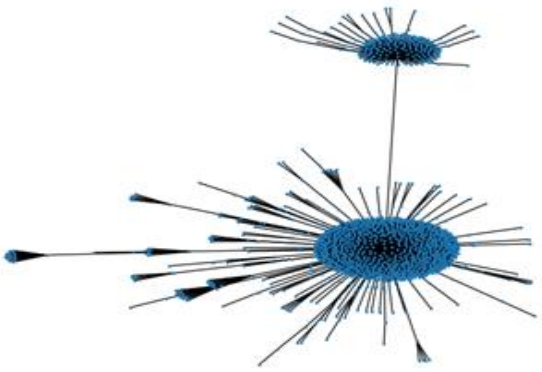

Figure 2. Cascade of a Misinformation Root Message

A cascade can be characterized by the number of steps (i.e., hops) the root message has traveled, or the times cascade nodes are posted. Propagation measurements for hop-based cascades often include size (number of users/posts involved), depth (maximum number of post hops from the root message, where a hop is a post by a new user), breadth (the maximum number of users involved in the cascade at any depth), and structural virality (average distance between all pairs of nodes in a cascade) of the cascade. For time-based cascades, the measures can be lifetime, real-time heat, and overall heat of a cascade (Kwon et al., 2013; Vosoughi et al., 2018; X. Zhou & Zafarani, 2018). Plenty of examples of cascades analysis can be found in literature, which help us understand the dynamics of misinformation propagation on social media (Bondielli & Marcelloni, 2019; X. Zhang & Ghorbani, 2020; X. Zhou & Zafarani, 2018). As one of the most representative examples, Vosoughi et al. (2018) analyzed multiple features of various types of root messages (e.g., real news, fake news, political, non-political etc.) and generated interesting observations on how fake news propagates on Twitter during the 2016 presidential election. This dissertation follows their methods to reconstruct and measure the cascades collected on Twitter.

2.3 Hypotheses and Research Model

Essay I mainly examines three sets of relationships: the one between producers and root messages, the one between producers and their consumers, and the one between root messages produced by these producers and their consumers. Furthermore, this study focuses on the Twitter platform and consumers. Therefore, the Root Message Profile is instantized as Root Message Profile, and Consumer Profile is instantized as Consumer Profile. Measures of these constructs and introduced below.

Prior work has identified a set of features of the root messages and users engaged in misinformation (e.g., Paschen, 2019; Rangel et al., 2020; Reis et al., 2019a, 2019b; X. Zhou & Zafarani, 2019). These features have been shown to be indicators or predictors of the veracity of root messages; they also indicate how virally misinformation root messages can propagate. These features include characteristics of social networks of producers and/or receivers (e.g., number followers and friends of producers or receivers), online behavior or them on the platform (e.g., how many posts have receivers made), basic text statistics of root messages (e.g., length of messages), text quality metrics of root messages (e.g., readability and subjectivity), and strength and types of sentiments and emotions expressed in root messages. Following prior studies on such as those listed above, I also examine the same pool of features of the producers, consumers, and root messages in relation to the constructs employed in Essay I in order to clarify these components' roles in misinformation production and diffusion.

Producer Profile is measured on producers of root messages by a set of variables that described the identify and behavior of the producers. All these variables are exposed by the Twitter platform. They are presented in Table 1.

Table 1. Measures of Producer Profile

Metainformation of Account		**Sentiment and Emotion of User Description**				**Text Statistics of User Description**	
Variable Name	**Data Type**	**Variable Name**	**Data Type**	**Variable Name**	**Data Type**	**Variable Name**	**Data Type**
Count of followers of the account	Integer	Score of emotion: fear	Float in [0, 1]	Score of sentiment: negative	Float in [0, 1]	Subjectivity score	Float in [0, 1]
Count of friends of the account	Integer	Score of emotion: anger	Float in [0, 1]	Score of sentiment: neutral	Float in [0, 1]	Readability score	Float in [0, 1]
Count of times the account being listed (similar to subscribed) by other users	Integer	Score of emotion: sadness	Float in [0, 1]	Score of sentiment: positive	Float in [0, 1]	Count of characters	Integer
Count of favorites (i.e., likes) received by the account	Integer	Score of emotion: disgust	Float in [0, 1]	Score of sentiment: compound	Float in [0, 1]	Count of words	Integer
Count of statuses (i.e., tweets) posted by the account	Integer	Score of emotion: surprise	Float in [0, 1]			Count of unique words	Integer
Whether the account as enabled to expose its geographical information in its online profile	Boolean	Score of emotion: anticipation	Float in [0, 1]			Count of sentences	Integer

Whether the user's identity has been verified by the platform	Boolean	Score of emotion: trust	Float in [0, 1]			Characters per word	Float > 0
		Score of emotion: joy	Float in [0, 1]			Words per sentence	Float > 0
		Score of emotion: any positive emotions	Float in [0, 1]				
		Score of emotion: any negative emotions	Float in [0, 1]				

Consumer Profile is initially measured using the individual variables present in Table 2. Each variable name in the table represents two variables—mean and media—measured on all the consumers who have consumed a given root message. These variables are further divided into four groups, each describing the consumers' activeness on Twitter, their situation of socialization, text quality of their user description, and sentiment and emotion expressed in their user description. Variables in each group is then normalized into [0, 1] and averaged into a single numeric score (following existing studies like this: Badaro et al., 2018). In turn, Consumer Profile is measured by four scores, including the *Activeness Score, Socialization Score, Text Quality Score, and Sentiment-Emotion Score*. The composing variables of each score is presented in Table 2.

Table 2. Measures of Consumer Profile

Activeness Score		**Socialization Score**		**Text Quality Score**		**Sentiment-Emotion Score**			
Variable Name	**Data Type**	**Variable Name**	**Data Type**	**Variable Name**	**Data Type**	**Variable Name**	**Data Type**	**Variable Name**	**Data Type**
Account age in days	Float > 0	Count of followers of the account	Float > 0	Score of readability	Float in [0, 1]	Score of emotion: fear	Float in [0, 1]	Score of emotion: any positive emotions	Float in [0, 1]
Count of favorites (i.e., likes) received by the account	Float > 0	Count of friends of the account	Float > 0	Score of subjectivity	Float in [0, 1]	Score of emotion: anger	Float in [0, 1]	Score of emotion: any negative emotions	Float in [0, 1]

Count of posts made by the account	Float > 0	Count of times the account being listed (similar to subscribed) by other users	Float > 0	Count of characters	Float > 0	Score of emotion: sadness	Float in [0, 1]	Score of sentiment: negative	Float in [0, 1]
		Percentage of users who enable to expose their geographical information in their online profile	Float in [0, 1]	Count of words	Float > 0	Score of emotion: disgust	Float in [0, 1]	Score of sentiment: neutral	Float in [0, 1]
		Percentage of users who are verified for their identify by platform	Float in [0, 1]	Count of unique words	Float > 0	Score of emotion: surprise	Float in [0, 1]	Score of sentiment: positive	Float in [0, 1]
		Percentage of users whose accounts are protected	Float in [0, 1]	Count of sentences	Float > 0	Score of emotion: anticipation	Float in [0, 1]	Score of sentiment: compound	Float in [0, 1]
				Characters per word	Float > 0	Score of emotion: trust	Float in [0, 1]		
				Words per sentence	Float > 0	Score of emotion: joy	Float in [0, 1]		
						Score of emotion: any positive emotions	Float in [0, 1]		
						Score of emotion: any negative emotions	Float in [0, 1]		

Root Message Profile is also measured using a larger set of variables exposed by Twitter initially. These variables are further divided into three groups, each describing the text quality of the root message, the sentiment and emotion expressed in the root message, and the propagation (e.g., how many retweets, and how many likes) of the root message. Similarly, each group of variables are normalized and then averaged into a single numeric score. Thus, Root Message Profile is measured by three scores, including the *Text Quality Score*, *Sentiment-Emotion Score*, and *Propagation Score*. The composing variables of each score is presented in Table 3 below.

Table 3. Measures of Root Message Profile

Text Quality Score		**Sentiment-Emotion Score**		**Sentiment-Emotion Score**		**Propagation Score**	
Variable Name	**Data Type**	**Variable Name**	**Data Type**	**Variable Name**	**Data Type**	**Variable Name**	**Data Type**
Readability score	Float in [0, 1]	Score of emotion: fear	Float in [0, 1]	Score of sentiment: negative	Float in [0, 1]	Cascade size (i.e., number of times the root message has been passed by)	Integer
Subjectivity score	Float in [0, 1]	Score of emotion: anger	Float in [0, 1]	Score of sentiment: neutral	Float in [0, 1]	Count of favorites (i.e., likes) received by the root messages	Integer
Count of characters	Integer	Score of emotion: sadness	Float in [0, 1]	Score of sentiment: positive	Float in [0, 1]		
Count of words	Integer	Score of emotion: disgust	Float in [0, 1]	Score of sentiment: compound	Float in [0, 1]		
Count of unique words	Integer	Score of emotion: surprise	Float in [0, 1]				
Count of sentences	Integer	Score of emotion: anticipation	Float in [0, 1]				
Characters per word	Float > 0	Score of emotion: trust	Float in [0, 1]				
Words per sentence	Float > 0	Score of emotion: joy	Float in [0, 1]				
Subjectivity score	Float in [0, 1]	Score of emotion: any positive emotions	Float in [0, 1]				
		Score of emotion: any negative emotions	Float in [0, 1]				

In this study, the unit of study is a root message, which is also the unit of observation. For instance, each root message is associated with a measurement of Activeness Score of its Consumer Profile. To further aggregate the individual scores under Root Message Profile, all the root messages collected are clustered over the three individual scores under Root Message Profile. Then, the cluster label computed for each root message becomes a single, categorical variable that simultaneously characterizes the text quality, sentiment and emotion, and propagation of the root message. I denote this variable *Root Message Profile Clustered*. Similarly, a cluster label is computed for each root message incorporating the Activeness Score, Socialization Score, Text Quality Score, and Sentiment-Emotion Score of the consumers of this root message. In this way, each root message has another single variable that simultaneously characterizes its consumers' activeness, socialization, text quality in user description, and sentiment and emotion in user description. I denote this variable *Consumer Profile Clustered*.

With all these variables and measures defined, I propose the following hypotheses, which are illustrated in the research model in Figure 3. First, I posit that depending on the characteristics of the producers (namely, their online profiles), they can produce root messages with different content, and the root messages they produce might propagate differently. Given that the content and propagation of root messages are both characterized by the online profile of root messages, I hypothesize that:

H1: Producer Profile influences: (H1a) the text quality; (H1b) sentiment and emotion; and (H1c) propagation of the root messages produced by these producers.

Furthermore, root message receivers with different characteristics (as characterized by their online profile) might make different decision on whether or not to consume (e.g., to retweet) the root messages depending on the characteristics (i.e., the online profile) of the producers of these root messages. Meanwhile, the root messages might impact receivers' decision too, and thus need to be accounted for (as control variables). Therefore, I hypothesized that:

H2: Producer Profile influences which users consume the root messages produced by these producers; these users are characterized/differentiated by: (H2a) their activeness on the

platform; (H2b) their activities of socialization; (H2c) the text quality of their user description; and (H2d) sentiment and emotion expressed in their user description.

Furthermore, characteristics of the body of root messages can influence how these root message propagate among receivers. Meanwhile, producers' online profile might be considered by the receivers when they decide whether to consume (i.e., propagate) the root messages and thus need to be accounted for (as control variables). Therefore, I hypothesized that:

H3: Linguistic characteristics of the body of root messages influence the propagation of these root messages.

Finally, root message receivers with different characteristics (as characterized by their online profile) might make different decision on whether or not to consume (e.g., to retweet) the root messages depending on the profile of the root messages. Meanwhile, the receivers might consider the producers' profile when they decide whether or not to consume the root messages. Thus, producers' profile needs to be accounted for (as control variables). Therefore, I hypothesized that:

H4: Root Message Profile influences which users consume these root messages; these users are characterized/differentiated by: (H4a) their activeness on the platform; (H4b) their activities of socialization; (H4c) the text quality of their user description; and (H4d) sentiment and emotion expressed in their user description.

H4a to H4d are measured using the individual scores under Consumer Profile separately.

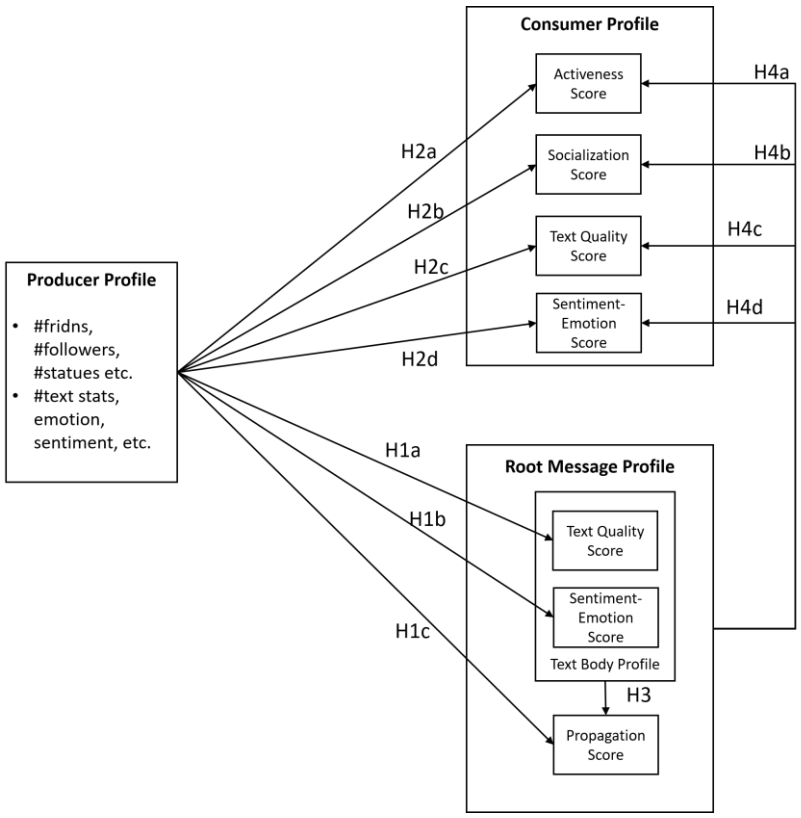

Figure 3. Research Model of Essay I

2.4 Methodology

2.4.1 Data

2.4.1.1 Data Description

The research essays have been conducted as empirical studies based on a collection of root messages from Twitter. The on-going process of data collection was started in early April 2020 and is expected to continue even beyond this study. Currently, more than 1300 root messages (namely, tweets produced by producers and propagated by consumers) together with their retweet cascades and replies have been collected. I have followed the method in prior studies (Goel et al., 2012; Vosoughi et al., 2018) to make sure that each of the collected root messages has been rated for its veracity by at least one of the four highly popular fact-checking websites, including truthorfiction, snopes, politifact, and factcheck. Meanwhile, all the root messages are classified as related to COVID-19 by these websites.

A subset of root messages in the data collection above were used in this study, which includes 628 pieces of root messages of misinformation (namely, received a label from the fact-

checking websites indicating it is misinformation, such as a label of “False”, “Not True”, “Pants on Fire”, etc.), 252 pieces of root messages identified as authentic information (namely, received a label indicating it is authentic, such as a label of “True”, “Mostly True”, etc.), and 47 root messages of neither misinformation nor authentic information (with a label of, e.g., “Unknown”, “Satire”, etc.). The data set used in this study is described in Table 4.

Table 4. Descriptive Statistics of Data

Veracity	Num. Root Messages	Num. Consumers per R. Message	Num. Likes per R. Message	Num. Producers	Num. Followers per Prod.	Num. Friends per Prod.	Num. Likes per Prod.	Num. Listed per Prod.	Num. Posts per Prod.
Misinformation	628	1,796	7,083	574	426,412	9,372	39,008	1,662	57,715
Authentic Info.	252	1,984	7,459	227	1,743,593	11,724	50,871	7,998	118,359
Neither (unknown, satire, etc.)	47	5,591	19,970	44	301,997	13,807	71,828	1,878	93,085

2.4.1.2 Data Preprocessing

The root messages (i.e., tweets identified by the fact-checking websites as containing misinformation related to COVID-19) and the retweets made for these root messages were downloaded via the Twitter API (Twitter, n.d.). This is a RESTful API exposed by the Twitter platform for the users to search for and collect tweets in batches.

Each root message and its retweets were stored in the plain-text JSON format in a separate file, with each JSON object containing all attributes describing a single tweet, such as the tweet’s creation time, the ID and user description of the user who posted it, the text body of the tweet, various metainformation of the tweet (e.g., count of retweets, replies, likes, etc.). All the variables of Producer Profile and Root Message Profile can be found or computed from the downloaded JSON objects.

Then, each root message and its retweets were arranged into a cascade, in which the retweets were ordered ascendingly by their creation time. At this point, the variables of Consumer Profile were computed for each root message over it retweets made within the first 48 hour since the creation of the root message. Next, a single table was created with each row representing all the variables measuring an individual root message. Finally, the veracity and the variables of Producer Communicative Intention and Production Approach were annotated and

stored as binary values in the table. The data preprocessing was implemented mainly in Python (v3.8.3),

2.4.2 Research Methods

To complete Essay I, an empirical study has been conducted using the Twitter data. That means, in this study the producers only refer to users of Twitter (called Twitters). Root messages only refer to posts made by the producers on Twitter, namely, tweets. In particular, I only focus on such consumers who are also Twitter users and have forwarded any root message without commenting, namely, who have retweeted the root message. That means each consumer has created a retweet to the root message he or she consumed.

Statistical analyses were conducted on the root messages with misinformation (n = 628) in order to verify or reject the hypotheses. A series of regression analysis addressing the hypotheses between the constructs. All statistical analyses were conducted at the level of the root messages. Linear regression with robust standard errors clustered on producers was employed. This method was selected due to the possibility that producers might influence the root messages and create cluster effects among the observations.

The data collection was performed using Python (v3.8). Data preprocessing such as data cleaning and feature generation was performed using Python and R (v4.0). The regression analysis in this essay was performed using the linear regression package with robust standard errors in R. The sentiment scores involved in measuring any construct in this dissertation were computed using the VADER toolkit in NLTK library (Steven Bird & Lingling Tan, n.d.) of Python based on the VADER (Valence Aware Dictionary and Sentiment Reasoner) sentiment lexicon (Hutto & Gilbert, 2014). The emotion scores were computed using the NRCLex library (Mark M. Bailey, n.d.) of Python based on the National Research Council Canada (NRC) affect lexicon (Mohammad et al., 2013).

2.5 Analyses and Results

2.5.1 Influence of Producer Profile on Root Message Profile (H1)

Analyses in this subsection are performed using linear regression. First, the linear model in Equation (1) is established and fitted on the collection of misinformation root messages:

$$TextQualityScore = ProducerProfile \quad (1)$$

TextQualityScore represents the Text Quality Score of Root Message Profile. *ProducerProfile* includes all the individual variables included in the construct Producer Profile and their order-

two interaction terms. The interaction terms were included in the regression model because inspecting the coefficients of these terms can indicate if any two of the individual variables present influences on the dependent variable collectively and if they have interactive effects (i.e., if one of these two measures affects the other's influence on the dependent variable) (Jaccard et al., 1990).

All abbreviations of variable names in Equation (1) are described in Table 5. Regression results indicate that Producer Profile is strongly associated with the Text Quality Score (H1a) and Sentiment-Emotion Score of Root Message Profile (H1b). In terms of the Text Quality Score (see Table 6), there are 8 interaction terms of the variables of Producer Profile that are significantly associated with this DV. Among them, the interaction terms involving certain negative emotion, such as prdPrf_sub * prdPrf_emoFea, prdPrf_emoFea * prdPrf_emoAng, and prdPrf_emoAng * prdPrf_emoDis, preset a strong, positive association with the Text Quality Score. Meanwhile, interaction terms involving the text complexity of producers' user description, such as prdPrf_emoSad * prdPrf_chaCnt and prdPrf_emoJoy * prdPrf_worSen, present a strong, negative association with the Text Quality Score.

Table 5. Abbreviations in Variable Names

Linguistic Characteristics				**Account Meta-information**	
Text Features		*Emotion Features (floating number in (0,1))*		*folCnt*	Count of followers of the account
rea	Readability of text instance (floating number in (0, 1))	*emoAng*	Anger	*friCnt*	Count of friends of the account
sub	Readability of text instance (floating number in (0, 1))	*emoDis*	Disgust	*lisCnt*	Count of times the account being listed (similar to subscribed) by other users
chaCnt	Count of characters	*emoFea*	Fear	*favCnt*	Count of favorites received by the account
worCnt	Count of words	*emoJoy*	Joy	*staCnt*	Count of statuses (i.e., tweets) posted by the account
uniWorCnt	Count of unique words	*emoSad*	Sadness		
senCnt	Count of sentences	*emoSur*	Surprise		
chaWor	Count of characters per word (floating number)	*emoTru*	Trust		
worSen	Count of words per sentence (floating number)	*emoNeg*	Overall negative emotion		
Sentiment Features (floating number in (0, 1))		*emoPos*	Overall positive emotion		
senNeg	Negative sentiment				
senNeu	Neutral sentiment				
senPos	Positive sentiment				
senCom	Composite score of sentiment				

Table 6. Regression of Producer Profile on Root Message Profile - Text Quality Score (H1a)

IV's	Producer Profile
DV	Root Message Profile – Text Quality Score
Num. Obs.	590
R2	0.727
Num. Sig. Main IV's	0
Num. Sig. Main IV's Inter.	8
Num. Sig. CV's	0
prdPrf_sub * prdPrf_emoFea	2.173+
prdPrf_emoFea * prdPrf_emoAng	7.09+++
prdPrf_emoAng * prdPrf_emoDis	10.832+++
prdPrf_emoSad * prdPrf_chaCnt	-0.1+++
prdPrf_emoJoy * prdPrf_worSen	-0.136+
prdPrf_chaWor * prdPrf_rea	-3e-04+++
prdPrf_favCnt * prdPrf_emoFea	-1e-05+++
prdPrf_staCnt * prdPrf_emoDis	-1e-05+++
p-value +: <0.05, ++: <0.01, +++: <0.001	

Then, the linear model in Equation (2) is established and fitted on the collection of misinformation root messages:

$$SentimentEmotionScore = ProducerProfile \quad (2)$$

SentimetEmotionScore represents the Sentiment-Emotion Score of Root Message Profile. *ProducerProfile* includes all the individual variables included in the construct Producer Profile and their order-two interaction terms. Results in Table 7 show that the association with the Sentiment-Emotion Score of Root Message Profile is even stronger. In total, there are two individual variables of Producer Profile significantly associated with this score: prdPrf_worSen presents a strong positive association, and prdPrf_worCnt presents a strong negative association. Furthermore, 38 interaction terms of variables of Producer Profile are detected significantly associated with the Sentiment-Emotion Score. For instance, interaction terms involving certain sentiment and word count project the strongest, positive association, such as prdPrf_senNeg * prdPrf_worCnt, prdPrf_senNeu * prdPrf_worCnt, and prdPrf_senPos * prdPrf_worCnt.

Table 7. Regression of Producer Profile on Root Message Profile – Sentiment-Emotion Score (H1b)

IV's	Producer Profile
DV	Root Message Profile – Sentiment-Emotion Score
Num. Obs.	591
R2	0.727
Num. Sig. Main IV's	2
Num. Sig. Main IV's Inter.	38
Num. Sig. CV's	0
prdPrf_worSen	25.443+++
prdPrf_worCnt	-169.947+++
prdPrf_senNeg * prdPrf_worCnt	170.026+++
prdPrf_senNeu * prdPrf_worCnt	169.959+++
prdPrf_senPos * prdPrf_worCnt	170.012+++
prdPrf_senCom * prdPrf_emoAng	2.386+
prdPrf_emoAng * prdPrf_emoTru	2.309+
prdPrf_lisCnt * prdPrf_emoAng	0.001+++
prdPrf_emoAng * prdPrf_chaWor	0.637+++
prdPrf_senNeg * prdPrf_worSen	-25.356+++
prdPrf_senNeu * prdPrf_worSen	-25.382+++
prdPrf_senPos * prdPrf_worSen	-25.419+++
prdPrf_emoFea * prdPrf_emoTru	-1.061+++
prdPrf_emoAng * prdPrf_emoPos	-2.184+
prdPrf_emoSur * prdPrf_chaCnt	-0.028+++
prdPrf_senCom * prdPrf_emoSur	-0.976+++
prdPrf_emoDis * prdPrf_worCnt	-0.307+
prdPrf_emoDis * prdPrf_chaWor	-1.065+
prdPrf_emoPos * prdPrf_uniWorCnt	-0.057+++
prdPrf_emoPos * prdPrf_chaWor	-0.053+++
prdPrf_emoNeg * prdPrf_chaCnt	-0.018+++
prdPrf_emoSad * prdPrf_chaCnt	-0.028+++
prdPrf_chaWor * prdPrf_worSen	-0.011+++
p-value +: <0.05, ++: <0.01, +++: <0.001	

In comparison, interaction terms involving sentiment and sentence complexity produce the strongest, negative association, such as prdPrf_senNeg * prdPrf_worSen, prdPrf_senNeu * prdPrf_worSen and prdPrf_senPos * prdPrf_worSen. Finally, interaction terms involving two different type emotion can also produce strong, negative association with the Sentient-Emotion Score, such as prdPrf_emoFea * prdPrf_emoTru and prdPrf_emoAng * prdPrf_emoPos.

Furthermore, the associations underlying H1c (pointing to the Propagation Score of Root Message Profile) is tested without receiving significant result. Therefore, H1c is believed to be not supported.

2.5.2 Influence of Producer Profile on Consumer Profile (H2)

First, the linear model in Equation (3) is established and fitted on the collection of misinformation root messages:

$$ActivenessScore = ProducerProfile + RootMessageProfile \quad (3)$$

ActivenessScore represents the Activeness Score of Consumer Profile. *ProducerProfile* includes all the individual variables included in the construct Producer Profile and their order-two interaction terms. *RootMessageProfile* represents the control variables, including all individual variables under the construct Root Message Profile.

We have observed substantial association between variables of Producer Profile and the Activeness Score (H2a) from the regression results (see Table 8). Two individual variables under Producer Profile, prdPrf_emoAng and prdPrf_emoJoy, present the strongest positive association with the Activeness Score. Further, 20 interaction terms of variables of Producer Profile are significantly associated with this score. Among them, interaction terms that produce the strongest positive impact are those that involve two different types of sentiments (e.g., prdPrf_senNeg * prdPrf_senPos). Strong negative influence are produced by interaction terms between the two types of highly influential emotions above and producer's sentiment, such as prdPrf_senNeg * prdPrf_emoAng, prdPrf_senNeg * prdPrf_emoJoy, prdPrf_senPos * prdPrf_emoAng and prdPrf_senPos * prdPrf_emoJoy. In addition, interaction terms between different types of emotions also result in negative influence, such as prdPrf_emoTru * prdPrf_emoNeg and prdPrf_emoPos * prdPrf_emoSad.

Table 8. Regression of Producer Profile on Consumer Profile - Activeness Score (H2a)

IV's	Producer Profile
DV	Consumer Profile - Active Score
Num. Obs.	590
R2	0.815
Num. Sig. Main IV's	2
Num. Sig. Main IV's Inter.	20
Num. Sig. CV's	0
prdPrf_emoAng	9052.93+++

prdPrf_emoJoy	4660.197+++
prdPrf_senNeg * prdPrf_senNeu	6.182+++
prdPrf_senNeg * prdPrf_senPos	14.656+++
prdPrf_emoAng * prdPrf_chaWor	1.503+++
prdPrf_senCom * prdPrf_chaCnt	0.024+++
prdPrf_sub * prdPrf_chaCnt	0.017+++
prdPrf_senNeg * prdPrf_emoAng	-9066.871+++
prdPrf_senNeg * prdPrf_emoJoy	-4661.006+++
prdPrf_senNeu * prdPrf_emoAng	-9061.838+++
prdPrf_senNeu * prdPrf_emoJoy	-4661+++
prdPrf_senPos * prdPrf_emoAng	-9059.307+++
prdPrf_senPos * prdPrf_emoJoy	-4660.642+++
prdPrf_emoFea * prdPrf_chaWor	-1.108+++
prdPrf_emoTru * prdPrf_emoNeg	-1.143+++
prdPrf_emoPos * prdPrf_emoSad	-3.52+++
prdPrf_sub * prdPrf_worCnt	-0.073+++
prdPrf_sub * prdPrf_chaWor	-0.328+
prdPrf_emoFea * prdPrf_rea	-0.018+++
prdPrf_emoSur * prdPrf_worSen	-0.094+++
p-value +: <0.05, ++: <0.01, +++: <0.001	

Then, the linear model in Equation (4) is established and fitted on the collection of misinformation root messages:

$$SocializationScore = ProducerProfile + RootMessageProfile \quad (4)$$

SocializationScore represents the Socialization Score of Consumer Profile. *ProducerProfile* includes all the individual variables included in the construct Producer Profile and their order-two interaction terms. *RootMessageProfile* represents the control variables, including all individual variables under the construct Root Message Profile.

As regression results in Table 9 demonstrate, Producer Profile's influence on the Socialization Score of Consumer Profile is limited, though still significant. For example, interaction terms between positive emotion expressed in user description and certain text metrics (e.g., with the unique word count as in prdPrf_emoPos * prdPrf_uniWorCnt) produce positive influence on the Socialization Score. In comparison, interaction terms between certain negative emotions (e.g., sadness and anger) and sentiment or with positive emotion (e.g., trust) present negative influence on the Socialization Score, such as with prdPrf_senCom * prdPrf_emoSad and prdPrf_emoAng * prdPrf_emoTru.

Table 9. Regression of Producer Profile on Consumer Profile - Socialization Score (H2b)

IV's	Producer Profile
DV	Consumer Profile - Socialization Score
Num. Obs.	590
R2	0.788
Num. Sig. Main IV's	0
Num. Sig. Main IV's Inter.	17
Num. Sig. CV's	1
prdPrf_emoPos * prdPrf_uniWorCnt	0.067+
prdPrf_senCom * prdPrf_emoSad	-1.281+++
prdPrf_emoAng * prdPrf_emoTru	-1.769+++
prdPrf_emoPos * prdPrf_worCnt	-0.055+
prdPrf_emoTru * prdPrf_uniWorCnt	-0.039+++
prdPrf_uniWorCnt * prdPrf_chaWor	0.019+++
prdPrf_emoSur * prdPrf_rea	-0.004+++
p-value +: <0.05, ++: <0.01, +++: <0.001	

Furthermore, the linear model in Equation (5) is established and fitted on the collection of misinformation root messages:

$$TextQualityScore = ProducerProfile + RootMessageProfile \quad (5)$$

TextQualityScore represents the Text Quality Score of Consumer Profile. *ProducerProfile* includes all the individual variables included in the construct Producer Profile and their order-two interaction terms. *RootMessageProfile* represents the control variables, including all individual variables under the construct Root Message Profile.

Results in Table 10 indicate that variables under Producer Profile are only weakly associated with the Text Quality Score (H2c). Significant but weak, positive influence are observed from interaction terms involving the favorite count and status count of the producer, such as prdPrf_favCnt * prdPrf_chaCnt, prdPrf_favCnt * prdPrf_worCnt, and prdPrf_staCnt * prdPrf_emoFea. Slightly stronger, negative influence is produced by the interaction term between the fear expressed and text readability of producers' user description (prdPrf_emoFea * prdPrf_rea).

Table 10. Regression of Producer Profile on Consumer Profile – Text Quality Score (H2c)

IV's	Producer Profile
DV	Consumer Profile - Text Quality Score
Num. Obs.	587
R2	0.745
Num. Sig. Main IV's	0
Num. Sig. Main IV's Inter.	4
Num. Sig. CV's	0
prdPrf_favCnt * prdPrf_chaCnt	0+++
prdPrf_favCnt * prdPrf_worCnt	0+++
prdPrf_staCnt * prdPrf_emoFea	-1e-05+++
prdPrf_emoFea * prdPrf_rea	-0.015+++
p-value +: <0.05, ++: <0.01, +++: <0.001	

Furthermore, the associations underlying H2d (pointing to the Sentiment-Emotion Score of Consumer Profile) is tested without receiving significant result. Therefore, H2d is believed to be not supported.

2.5.3 Influence of Root Messages' Text Body on Root Messages' Propagation (H3)

The linear model in Equation (6) is established and fitted on the collection of misinformation root messages:

$$PropagtionScore = RootMessageTextBody + ProducerProfile \quad (6)$$

PropagationScore represents the Propagation Score of Root Message Profile. *RootMessageTextBody* include all the individual variables associated the text body of root messages (i.e., variables associated with the Text Quality Score and Sentiment-Emotion Score of Root Message Profile) and their order-two interaction terms. *ProducerProfile* includes all the individual variables under the construct Producer Profile, which is the set of control variables.

As the results in Table 11 indicate, significant association has been detected within the Root Message Profile between the variables associated with root messages' text body and the Propagation Score (H3). There are six interaction terms among the text body variables associated with the Propagation score significantly. For example, the interaction term between anger and surprise in the text body (rooPrf_emoAng * rooPrf_emoSur) is strongly and positively associated with the score. Further, multiple interaction terms involving joy and other types of positive emotions (e.g., rooPrf_emoTru * rooPrf_emoJoy, rooPrf_emoSur * rooPrf_emoJoy, and rooPrf_emoPos * rooPrf_emoJoy) project strong, negative influence upon the score.

Table 11. Regression of Root Message Profile – Text Body on Root Message Profile - Propagation Score (H3)

IV's	Root Message Profile – Text Body Var.
DV	Root Message Profile - Propagation Score
Num. Obs.	590
R2	0.535
Num. Sig. Main IV's	0
Num. Sig. Main IV's Inter.	6
Num. Sig. CV's	2
rooPrf_emoAng * rooPrf_emoSur	2.116+
rooPrf_emoTru * rooPrf_emoJoy	-1.217+++
rooPrf_emoSur * rooPrf_emoJoy	-1.69+++
rooPrf_emoPos * rooPrf_emoJoy	-1.355+++
rooPrf_emoDis * rooPrf_rea	-0.008+++
rooPrf_worSen * rooPrf_rea	3e-04+++
p-value +: <0.05, ++: <0.01, +++: <0.001	

2.5.4 Influence of Root Message Profile on Consumer Profile (H4)

First, the linear model in Equation (7) is established and fitted on the collection of misinformation root messages:

$$ActivenessScore = RootMessageProfile + ProducerProfile \quad (7)$$

ActivenessScore represents the Activeness Score of Consumer Profile. *RootMessageProfile* includes all the individual variables of the construct Root Message Profile and their order-two interaction terms. *ProducerProfile* includes all the individual variables under the construct Producer Profile, which is the set of control variables.

We can see from the results that there are three interaction terms significantly associated with the Activeness Score (see Table 12). Strong, negative influence is detected from interaction terms between the unique word count and various emotion in the text body of the root messages, including rooPrf_emoPos * rooPrf_uniWorCnt and rooPrf_emoDis * rooPrf_uniWorCnt.

Table 12. Regression of Root Message Profile on Consumer Profile - Activeness Score (H4a)

IV's	Root Message Profile
DV	Consumer Profile - Activeness Score
Num. Obs.	590
R2	0.536
Num. Sig. Main IV's	0

Num. Sig. Main IV's Inter.	3
Num. Sig. CV's	2
rooPrf_favCnt * rooPrf_emoJoy	-1e-04+++
rooPrf_emoPos * rooPrf_uniWorCnt	-0.038+++
rooPrf_emoDis * rooPrf_uniWorCnt	-0.108+++
p-value +: <0.05, ++: <0.01, +++: <0.001	

Then, the linear model in Equation (8) is established and fitted on the collection of misinformation root messages:

$$SocializationScore = RootMessageProfile + ProducerProfile \quad (8)$$

SocializationScore represents the Socialization Score of Consumer Profile. *RootMessageProfile* includes all the individual variables of the construct Root Message Profile and their order-two interaction terms. *ProducerProfile* includes all the individual variables under the construct Producer Profile, which is the set of control variables.

Eight interaction terms are found associated with the Socialization Score significantly (see Table 13). Among them, three interaction terms involving certain negative emotions create the strongest positive influence on the Socialization Score, including rooPrf_emoAng * rooPrf_emoDis, rooPrf_emoSad * rooPrf_emoDis, and rooPrf_emoSad * rooPrf_uniWorCnt. By contrust, the interaction term between trust expressed in text body and the word count of sentences (rooPrf_emoTru * rooPrf_worSen) projects significant but weak negative influence on the Socialization Score.

Table 13. Regression of Root Message Profile on Consumer Profile - Socialization Score (H4b)

IV's	Root Message Profile
DV	Consumer Profile - Socialization Score
Num. Obs.	590
R2	0.577
Num. Sig. Main IV's	0
Num. Sig. Main IV's Inter.	8
Num. Sig. CV's	5
rooPrf_emoAng * rooPrf_emoDis	2.345+++
rooPrf_emoSad * rooPrf_emoDis	2.407+++
rooPrf_emoSad * rooPrf_uniWorCnt	0.063+
rooPrf_emoSur * rooPrf_rea	0.003+++
rooPrf_emoPos * rooPrf_rea	0.002+++
rooPrf_emoTru * rooPrf_worSen	-0.033+

rooPrf_worCnt * rooPrf_rea	2e-04+++
rooPrf_senCnt * rooPrf_rea	-3e-04+++
p-value +: <0.05, ++: <0.01, +++: <0.001	

Next, the linear model in Equation (9) is established and fitted on the collection of misinformation root messages:

$$TextQualityScore = RootMessageProfile + ProducerProfile \quad (9)$$

TextQualityScore represents the Text Quality Score of Consumer Profile. *RootMessageProfile* includes all the individual variables of the construct Root Message Profile and their order-two interaction terms. *ProducerProfile* includes all the individual variables under the construct Producer Profile, which is the set of control variables.

For the Text Quality Score, one individual variable and 20 interaction terms are found significant (see Table 14). The sentence count (rooPrf_senCnt) alone projects strong, negative influence on the Text Quality Score. However, when this variable interacts with strong sentiment (e.g., in rooPrf_senNeg * rooPrf_senCnt and rooPrf_senPos * rooPrf_senCnt), the influence is revered to be positive. Multiple interaction terms of emotion involving anger and sadness also create strong, positive influence on this score, such as rooPrf_emoAng * rooPrf_emoNeg, rooPrf_emoAng * rooPrf_emoSad, and rooPrf_emoTru * rooPrf_emoSad. In contrast, interaction terms between a positive and a negative sentiment or emotion produce strong, negative influence on the Text Quality Score, such as rooPrf_emoTru * rooPrf_emoDis, rooPrf_senCom * rooPrf_emoAng, rooPrf_emoFea * rooPrf_emoTru and rooPrf_emoFea * rooPrf_emoPos. Moreover, interaction terms involving certain text quality metrics (e.g., sentence count and words per sentence) and emotion expressed in the text body also produce considerable negative influence on the Text Quality Score, such as rooPrf_emoDis * rooPrf_senCnt, rooPrf_emoDis * rooPrf_worSen, and rooPrf_emoJoy * rooPrf_senCnt.

Table 14. Regression of Root Message Profile on Consumer Profile – Text Quality Score (H4c)

IV's	Root Message Profile
DV	Consumer Profile - Text Quality Score
Num. Obs.	587
R2	0.581
Num. Sig. Main IV's	1
Num. Sig. Main IV's Inter.	20
Num. Sig. CV's	2

rooPrf_senCnt	-34.841+++
rooPrf_senNeg * rooPrf_senCnt	34.653+++
rooPrf_senNeu * rooPrf_senCnt	34.756+++
rooPrf_senPos * rooPrf_senCnt	34.899+++
rooPrf_emoAng * rooPrf_emoNeg	2.434+++
rooPrf_emoAng * rooPrf_emoSad	1.819+++
rooPrf_emoTru * rooPrf_emoSad	1.78+
rooPrf_emoTru * rooPrf_emoDis	-2.668+++
rooPrf_senCom * rooPrf_emoAng	-1.078+
rooPrf_emoFea * rooPrf_emoTru	-1.254+
rooPrf_emoFea * rooPrf_emoPos	-1.234+
rooPrf_emoDis * rooPrf_senCnt	-0.236+++
rooPrf_emoDis * rooPrf_worSen	-0.111+++
rooPrf_emoJoy * rooPrf_senCnt	-0.19+++
rooPrf_senCom * rooPrf_worSen	-0.029+++
rooPrf_sub * rooPrf_worCnt	-0.023+++
p-value +: <0.05, ++: <0.01, +++: <0.001	

Finally, the linear model in Equation (10) is established and fitted on the collection of misinformation root messages:

$$SentimentEmotionScore = RootMessageProfile + ProducerProfile \quad (10)$$

SentimentEmotionScore represents the Sentiment-Emotion Score of Consumer Profile. *RootMessageProfile* includes all the individual variables of the construct Root Message Profile and their order-two interaction terms. *ProducerProfile* includes all the individual variables under the construct Producer Profile, which is the set of control variables.

The Sentiment-Emotion Score is found influenced by 23 interaction terms of variables under Root Message Profile (see Table 15). Among them, interaction terms between sadness and another type of emotion in the text body of root messages produce the strongest positive influence on the Sentiment-Emotion Score, such as rooPrf_emoSur * rooPrf_emoSad, rooPrf_emoNeg * rooPrf_emoSad, rooPrf_emoPos * rooPrf_emoSad, and rooPrf_emoTru * rooPrf_emoSad. Interaction terms between fear and some text metrics (e.g., character per word and word count) also project strong positive influence on this score, such as rooPrf_emoFea * rooPrf_chaWor and rooPrf_emoFea * rooPrf_worCnt. By contrast, interaction terms involving disgust or fear and another emotion can produce strong negative influence on the score, such as with rooPrf_emoSur * rooPrf_emoDis and rooPrf_emoTru * rooPrf_emoDis, and

rooPrf_emoFea * rooPrf_emoTru, rooPrf_emoFea * rooPrf_emoPos, and rooPrf_emoFea * rooPrf_emoNeg.

Table 15. Regression of Root Message Profile on Consumer Profile - Sentiment-Emotion Score (H4d)

IV's	Root Message Profile
DV	Consumer Profile - Sentiment-Emotion Score
Num. Obs.	590
R2	0.590
Num. Sig. Main IV's	0
Num. Sig. Main IV's Inter.	23
Num. Sig. CV's	3
rooPrf_emoSur * rooPrf_emoSad	1.526+++
rooPrf_emoNeg * rooPrf_emoSad	1.4+
rooPrf_emoPos * rooPrf_emoSad	1.124+
rooPrf_emoTru * rooPrf_emoSad	1.287+
rooPrf_emoSur * rooPrf_emoNeg	0.559+++
rooPrf_emoFea * rooPrf_chaWor	0.24+++
rooPrf_sub * rooPrf_worSen	0.012+++
rooPrf_emoFea * rooPrf_worCnt	0.056+++
rooPrf_emoFea * rooPrf_rea	0.004+++
rooPrf_emoSur * rooPrf_emoDis	-3.227+
rooPrf_emoTru * rooPrf_emoDis	-1.675+++
rooPrf_emoFea * rooPrf_emoTru	-0.84+
rooPrf_emoFea * rooPrf_emoPos	-0.518+++
rooPrf_emoFea * rooPrf_emoNeg	-0.551+++
rooPrf_emoDis * rooPrf_senCnt	-0.147+++
rooPrf_emoDis * rooPrf_worSen	-0.078+++
rooPrf_emoJoy * rooPrf_worSen	-0.065+++
rooPrf_emoPos * rooPrf_worSen	-0.023+++
rooPrf_senCom * rooPrf_senCnt	-0.044+
rooPrf_senCom * rooPrf_worSen	-0.024+
rooPrf_sub * rooPrf_worCnt	-0.015+++
rooPrf_emoSur * rooPrf_worSen	-0.018+++
p-value +: <0.05, ++: <0.01, +++: <0.001	

2.6 Discussion

In summary, most of the hypotheses in Essay I have been verified and supported by the analytic results. First, the observations have verified H1a and H1b. Overall, in the context of H1a, it is observed that individual measures of producer profile might not impact the text quality of root messages' profile. However, the interaction among these measures do impact the text quality of root messages. Specifically, producers who express negative emotion (such as fear (prdPrf_emoFea), anger (prdPrf_emoAng), or disgust (prdPrf_emoDis)) in their user description tend to produce root messages with high text quality (represented by high Text Quality Score); those who express strong emotion (such as sadness (prdPrf_emoSad) or joy (prdPrf_emoJoy)) with complex text (e.g., using long text (high prdPrf_chaCnt) or long sentence (high prdPrf_worSen)) tend to produce low-quality root messages. The findings above indicate that producers not only determine the characteristics of root messages explicitly by composing them (e.g., by choosing specific words and images). Some properties carried by the producers might manifest in root messages automatically with or without the producers' awareness. The sentiment, emotion, and text quality are examples of such properties of the producers. The online profiles of producers and root messages can be used to monitor the influence of producers effectively.

Regarding the sentiment and emotion of root messages in the context of H1b, producers who use long sentences (namely, high in prdPrf_worSen) or who express sentiment (high in prdPrf_senNeg or prdPrf_senPos) with long text (prdPrf_worCnt) in their user description tend to produce root messages that are more sentimental and emotional (with high Sentiment-Emotion Score). However, once they use long sentences (high in prdPrf_worSen) to express sentiment in user description or when they express complicated emotion (indicated by, e.g., the cooccurrence of fear (prdPrf_emoFea), trust (prdPrf_emoTru), and anger (prdPrf_emoAng)), they are prone to producing less sentimental and emotional root messages. These findings reflect that generally sentimental or emotional producers tend to express their sentimentality and emotionality in the root messages they produce. However, this link is more significant among producers who are able to achieve certain level of text quality.

H2a has been verified given the strong and multi-fold association between variables under Producer Profile and the Activeness Score of Consumer Profile. The results suggest that producers who express anger (measured by prdPrf_emoAng) or joy (measured by

prdPrf_emoJoy) in their user description tend to attract consumers who are highly active on Twitter as indicated by their higher Activeness Score. Interestingly, these two emotions can also help attract consumer who are inactive on Twitter, if these emotions are expressed with strong sentiment in user description (namely, high in prdPrf_senPos or prdPrf_senNeg). Further, some other types of emotions in user description, such as trust (prdPrf_emoTru) and sadness (prdPrf_emoSad), can also help producers to attract inactive consumers. The findings above suggest that misinformation receivers who are active and inactive on a platform might make different decision on whether or not to consume (retweet in this study) a given root message depending on the characteristics of the producer of this root message. In particular, active and inactive receivers might show different reaction to producers' sentimentality and emotionality in general. In turn, active users and inactive users tend to have different preference to producers and the root messages they produce regarding producers' sentimentality and emotionality. However, different emotions and sentiments expressed by the producers might have different influences on these receivers' preference.

Furthermore, H2b has also been verified based on the influence from multiple variables under Producer Profile upon the Socialization Score of Consumer Profile. Results of analyses suggest that producers can attract consumers who are more socialized on Twitter (with higher Socialization Score) if these producers express positive emotion (prdPrf_emoPos) with text of higher quality—characterized by a higher unique word count (prdPrf_uniWorCnt)—in their user description. By contrast, if producers express anger (prdPrf_emoAng) and trust (prdPrf_emoTru) simultaneously or sadness (prdPrf_emoSad) and more positive sentiment (prdPrf_senCom) at the same time, consumers who are less socialized tend to be attracted. The findings above suggest that misinformation receivers who are more socialized and less socialized on a platform might make different decision on whether or not to consume (retweet in this study) a given root message depending on the characteristics of the producer of this root message. Specifically, socialized and unsocialized receivers might present distinct reaction to producers' emotionality in general. In turn, socialized users and unsocialized users tend to have different preference to producers and the root messages they produce depending on producers' emotionality. Furthermore, the text quality producers can achieve might impact the influence of these producers' emotionality.

In support of H3, the results discussed in previous subsections suggest that anger (rooPrf_emoAng) and surprise (rooPrf_emoSur) expressed together in the text body can substantially promote the propagation of root messages. Meanwhile, if the text body contains much positive emotion simultaneously, especially joy (rooPrf_emoJoy) and trust (rooPrf_emoTru), propagation of the root messages can be impeded. These findings suggest that the textual content of the root messages can strongly impact the propagation (e.g., retweeted and liked for how many times) of these root messages. Emotionality of the textual content is among the key factors that determine root messages' propagation.

Furthermore, H4a is moderately supported based on the significant association between multiple variables under Root Message Profile and the Activeness Score of Consumer Profile. Results suggest that, for example, root messages simultaneously with strong emotion (e.g., disgust (rooPrf_emoDis)) and more complicated text (e.g., with more unique words (rooPrf_uniWorSen)) can attract consumers who are less active (with lower Activeness Score) on Twitter. These findings suggest that misinformation receivers who are active and inactive on a platform might make different decision on whether or not to consume (retweet in this study) a given root message depending on the characteristics of this root message. In particular, active and inactive receivers might present distinct reaction to root messages' emotionality. In turn, active receivers and inactive receivers tend to have different preference to root messages in terms of root messages' emotionality. Furthermore, text quality or complexity of the root messages can amplify the influence of root messages' emotionality on a certain type of receivers.

H4b is also supported by the results. Results suggest that, for example, root messages expressing multiple types of negative emotions (e.g., anger (rooPrf_emoAng), sadness (rooPrf_emoSad), and disgust (rooPrf_emoDis)) can attracted more socialized consumers (with higher Socialization Score). Differently, root messages expressing trust (high in rooPrf_emoTru) using longer sentences (high in rooPrf_worSen) tends to attract less socialized consumers. Such findings suggest that misinformation receivers who are more socialized and less socialized on a platform might make different decision on whether or not to consume (retweet in this study) a given root message depending on the characteristics of this root message. In particular, socialized and unsocialized receivers might present distinct reaction to root messages' emotionality. In turn, socialized receivers and unsocialized receivers tend to have different preference to root messages in terms of root messages' emotionality. Furthermore, text quality or

complexity of the root messages can affect the influence of root messages' emotionality on a certain type of receivers.

Further, H4c is verified based on the observations. Results suggest that longer root messages composed with more sentences (high in rooPrf_senCnt) tend to attract consumers who present low text quality in their user description (with low Text Quality Score). However, if such long text is combined with strong sentiment (e.g., high in rooPrf_senNeg or rooPrf_senPos), the root messages tend to attract consumers with high text quality in their user description. Second, root messages expressing negative emotion (such as anger (rooPrf_emoAng) or sadness (rooPrf_emoSad)) tend to be highly attractive to consumers showing higher text quality. In contrast, root message expressing a positive and a negative emotion simultaneously (such as trust vs disgust and fear vs trust) tend to attract consumers presenting lower text quality in their user description. Finally, root messages expressing strong emotion (such as disgust (rooPrf_emoDis) or joy (rooPrf_emoJoy)) by complicated text, which is characterized by using more sentences (high in sentence count (rooPrf_senCnt)) and longer sentences (high in words per sentence (rooPrf_worSen)) tend to attract consumers showing lower text quality in their user description.

The findings related to H4c suggest that misinformation receivers who present high text quality and low text quality in their user description on a platform might make different decision on whether or not to consume (retweet in this study) a given root message depending on the characteristics of this root message. In particular, receivers presenting differential text quality might show distinct reaction to root messages' text quality, sentimentality, and emotionality. In turn, high- and low-text quality receivers tend to have different preference to root messages in terms of root messages' text quality, sentimentality, and emotionality. Furthermore, any one of these three aspects of the root messages can affect the influences from the other two aspects of root messages on a certain type of receivers.

Finally, H4d is also supported by the results. Results of analyses suggest that root messages combing sadness (measured by rooPrf_emoSad) and another type of emotion can strongly attract consumers who express strong emotion and sentiment in their user description. Meanwhile, root messages expressing fear (rooPrf_emoFea) using complicated text (e.g., with longer words (rooPrf_chaWor) or more words (rooPrf_worCnt)) tend to be attractive for consumers with high emotion and sentiment. In comparison, root messages expressing fear

(rooPrf_emoFea) or disgust (rooPrf_emoDis) together with another type of emotion can attract consumers who are less sentimental and emotional in their user description.

The findings related to H4d suggest that misinformation receivers who are more emotional and less emotional in their user description on a platform might make different decision on whether or not to consume (retweet in this study) a given root message depending on the profile of this root message. In particular, receivers presenting differential emotionality might show distinct reaction to root messages' emotionality. In turn, more emotional and less emotional receivers tend to have different preference to root messages with respect to root messages' emotionality. Furthermore, text quality or complexity of the root messages can affect the influence of root messages' emotionality on particular types of receivers.

2.7 Conclusion

Essay I is engaged in clarifying the relationship among the three components of the production and diffusion of misinformation on social media—the producers, root messages, and consumers. The characteristics of these components manifest in their online profiles exposed by the platform. Taking the Twitter platform and the retweeting behavior as the testbench, this essay hypotheses several sets of relationships. The analytic results drawn from a large, realistic data set have supported most of the hypothesized relationships.

How the hypotheses are supported by the analytic results are summarized in Figure 4. First of all, the type and degree of emotion (especially negative emotion) and sentiment expressed by the producers in their user description and the text complexity (characterized by the length of text and length of sentences) of the producers' user description can influence the text quality and the amount of sentiment and emotion of the root messages produced by these producers (H1a and H1b). Furthermore, emotion (e.g., joy, trust, anger, and surprise) expressed in the text body of the root messages can influence the propagation of these root messages (H3). Moving on, the emotion, sentiment, and text quality of the root messages can further influence which users will retweet the root messages: are these users relatively more active and socialized on Twitter (H4a and H4b)? Are they good at writing (H4c)? Are they emotional and sentimental as reflected in their user description (H4d)?

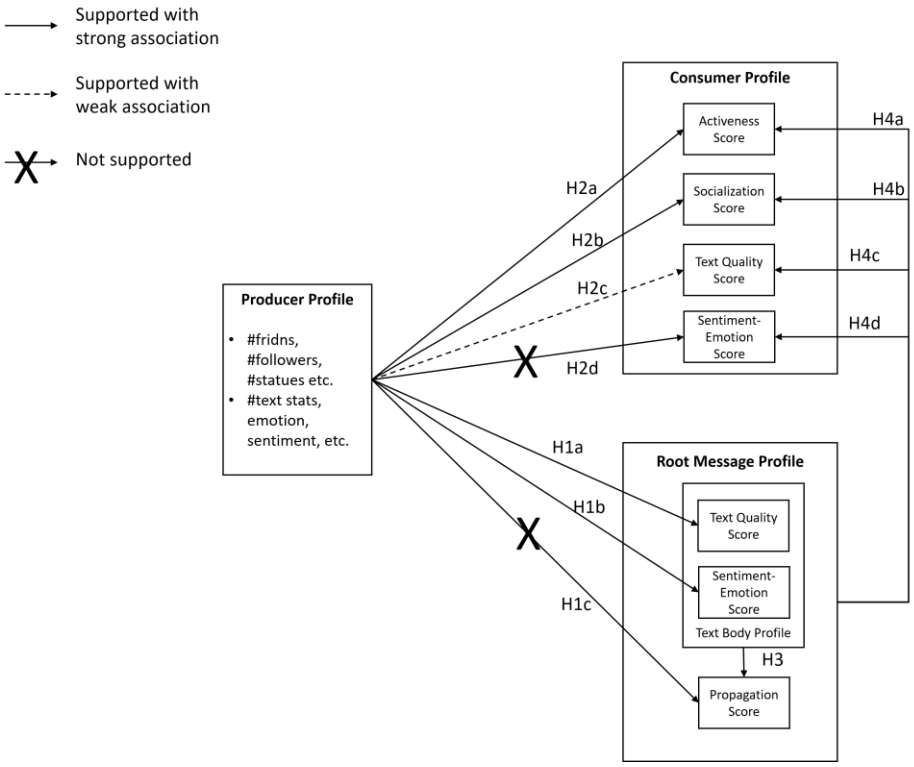

Figure 4. Support to Hypotheses in Essay I

On the other hand, producers' sentiment and emotion (e.g., the expression of anger, joy, sadness, and trust) expressed in their user description and the text quality (characterized by the length of text) of their user description can directly influence which types of users would retweet the root messages referring to how active and socialized these users are on Twitter (H2a and H2b).

In summary, the aforementioned findings fill the first gap in research as identified in the beginning of this dissertation that the basic mechanism that drives the diffusion of misinformation has not been revealed. Findings in Essay I indicate a set of features of producers that determine if the root messages they produce can propagate widely (i.e., gain large numbers of sharings and likes). Essay I also recognizes multiple features of misinformation root messages that can determine if receivers of root messages would consume (e.g., by retweeting) this piece of root message, rather than piece of root message.

Essay I also fills the second research gap that the producers and their roles in misinformation production an diffusion have not been clarified. Essay I recognizes many features in producers' online profile and examines their influences on the root messages and

consumers systematically. The findings produce a clearer and more comprehensive picture of the misinformation producers based on their online profiles.

CHAPTER III: ESSAY II: UNDERSTANDING THE IMPACT OF PRODUCER COMMUNICATIVE INTENTION, OPINION POLARITY, AND PRODUCTION APPROACH ON PRODUCTION AND DIFFUSION OF MISINFORMATION

3.1 Introduction

Essay II aims at filling the second, third and fourth research gaps that are specifically related to the producers of misinformation. In particular, the third and fourth gaps are addressed with the focus on the communicative intentions and opinion polarity of the producers, namely *why* they produce the messages, as well as the approaches they take to producing misinformation, namely *how* they produce misinformation. This essay examines the roles of these two categories of variables in the entire misinformation diffusion, i.e., how they influence (1) the content and propagation of root messages, and (2) who will consume these root messages.

Essays II extends the model of misinformation diffusion (see Figure 1): in the upstream lies the producer of misinformation. As a human, the producer holds her opinion, which might be polarized, on a specific issue. Also, communicative intentions form in her mind, which drive her to communicate some idea to her followers online; these intentions are conveyed by a set of speech acts she chooses to use. During composition of the root message, the producer performs (intentionally or unintentionally) some approaches and, as a result, a root message containing misinformation is produced.

Based on the model above, I posit that opinion polarity, communicative intentions of producers—represented by their use of speech acts—and the approaches they take to producing misinformation together act as the drivers of misinformation diffusion. Producers' intentions and approaches impact how root messages will spread (represented by propagation of root messages), and who will be interested in these messages (represented by the persona of consumers). In the remaining of the subsection, the key constructs supporting the central propositions are discussed in more detail.

Producer Communicative Intentions

This construct represents the communicative intentions of producers. A communication intention is defined as an intention of the speaker which is expected to be recognized by the receiver during communication. It declares what information is to be expressed (Melinger &

Levelt, 2004). Communicative intention of the message producer is the motive of the message, which drives the receiver's response (Hellbernd & Sammler, 2016).

Producers' communicative intentions are conveyed by the speech acts they choose (Searle & Searle, 1969; Villarroel Ordenes et al., 2017; R. Zhang et al., 2011). Speech acts are defined as linguistic acts performed in the process of speaking (Ilyas & Khushi, 2012; Sadock, 2004), which can be used to conceptualize speech as action (Vosoughi & Roy, 2016). Examples of speech acts include expressing our emotions, starting an argument, or insulting someone (Ilyas & Khushi, 2012). People construct their conversations through a set of speech acts to yield a specific communicative intention (Ebert et al., 2018). Performing a speech act is essentially expressing a certain intention in uttering certain words. A producer's communicative intention determines what speech acts she would perform (Sbisà, 2001).

Speech acts are measurable from posts on social media (e.g., Carr et al. 2012; Ilyas and Khushi 2012; Vosoughi and Roy 2016). Due to this measurability and the close association between communicative intentions and speech acts, Essay II uses speech acts as the measure of producers' communicative intentions on social media. The measuring process relies on the classic taxonomy of speech acts by Searle (Searle, 1976) described in Section 3.3.

Message Production Approaches

Message production approaches represent the approaches taken by producers to produce misinformation. Research has discovered a variety of approaches taken by the producers to create misinformation. Misinformation on social media can be created based on sensational events or public issues (Ahmed et al., 2020; Allcott & Gentzkow, 2017; Chiou & Tucker, 2018). During the production of misinformation, producers can amplify or frame (i.e. delimit the scope of) the events or issues strategically in order to interpret the situation in favor of their purposes (Marwick & Lewis, 2017). Variants of the same story can be created and delivered to social media in order to create memes, which can be passed along from person to person, growing into social phenomena (Qazvinian et al., 2011). To make their messages maximally sensational, multiple producers can leverage the participatory culture of social media and their social networks to jointly compose their stories (Jenkins, 2009). Technically, misinformation can be presented in various forms, such as posts, videos, blogs, pictures, or news websites (Bradshaw & Howard, 2018).

It can be seen from the discussion above that existing works tend to identify approaches to misinformation production by considering lots of information not available in the content of the root messages. This creates difficulties in measuring the approaches in practice. To this end, I propose to analyze misinformation production as behavior of interpersonal deception between producers and consumers, and use the Interpersonal Deception Theory (IDT) to model Production Approaches, namely, the approaches taken by producers during misinformation production (Buller & Burgoon, 1996). According to IDT, there are six semantic-syntactic approaches that can be taken to transform a piece of true information into misinformation (Burgoon et al., 1996). These approaches can be recognized through, and thus measured from, the linguistic characteristics of the produced messages (L. Zhou et al., 2004). It has been shown that linguistic characteristics of misinformation messages can indicate how the messages propagate and be accepted by receivers (Shu, Wang, et al., 2017). Hence, with the help of IDT which illustrates the linguistic characteristics of misinformation messages, the approaches taken by producers can more easily be related to Root Message Propagation and Consumer Profile, which are both important constructs of this study.

Producer Opinion Polarity

So far, research in opinion polarization on social media has been mainly confined in the political context. Overall, it is observed that networks of political re-posts (e.g., retweets) can exhibit a clearly segregated partisan structure, with low connectivity between politically liberal and conservative users (Conover et al., 2011). Some studies believe that the use of social media plays a role in reducing political polarization. For example, engagement on social media has been found to increase the network heterogeneity of partisan users (Lee et al., 2014). However, social media platforms are also found to facilitate exposure of novel information to users (Barberá, 2014). Some studies argue that, when such novel information is opposite to users' original political opinion, the degree of polarization among users might be increased (Bail et al., 2018). For instance, it is shown that republicans who followed a liberal Twitter bot became significantly more conservative; a similar polarization process also occurred with democrats who followed conservative bots (Bail et al., 2018).

There is also a significant body of research discussing polarization in the context of misinformation. However, this track of research mainly concentrates on the consumer side of misinformation, treating social media users as passive receivers of misinformation rather than its

producers. Studies as such have shown that users in polarized networks are more prone to diffusing misinformation (Bessi et al., 2015). This is mainly because polarized users tend to believe misinformation, as their vulnerability to misinformation is increased by directionally motivated reasoning—a cognitive process which occurs when one's belief is biased due to opinion polarization (Tucker et al., 2018). In turn, opinion polarization is considered a key indicator to recognize communities where misinformation is most likely to spread from users to users (Bessi et al., 2015). Following this line, users' degree of polarization is also used to predict potential topics of fake news (Vicario et al., 2019).

The observations above suggest that there is still insufficiency in literature in terms of: (1) exploring opinion polarization on social media in the non-political context; (2) investigating the effects of polarization on producers and their activity of misinformation production, and examining how these effects might further impact the other components of misinformation diffusion; and (3) examining the interaction between group opinion polarization and opinions of individuals.

Based on the constructs introduced above, several research questions are proposed in order to investigate how producers—characterized by their communicative intentions and their choice of approaches to misinformation production—would impact (1) propagation of the root messages, and (2) what types of users characterized by their persona will consume the root messages. These research questions are as follows.

Q1: How does the producers' opinion polarity influence their intention to produce misinformation?

Q2: How does the producers' intention of producing misinformation influence the approaches they take to producing this misinformation?

Q3: How do the communicative intentions held by producers when they produce root messages influence the content and propagation of the root messages produced by these producers?

Q4: How does the communicative intention held by producers influence which users consume the root messages produced by these producers?

Q5: How do the approaches chosen by producers when they produce root messages influence the content and propagation of the root messages produced by these producers?

Q6: How do the approaches chosen by producers influence which users consume the root messages produced by these producers?

To answer the questions above, I used the same data set described in Essay I and conducted statistical analyses on the root messages that labeled as misinformation by the fact checkers.

3.2 Theoretical Background and Literature Review

3.2.1 Producer Communicative Intentions and Speech Acts

3.2.1.1 Communicative Intentions of Producers

In Essay II, the construct Producer Communicative Intentions refers to the communicative intensions held by producers during misinformation production. This essay examines the role of producers' communicative intentions in misinformation diffusion: how these intentions expressed in root messages influence (1) the way in which root messages propagate, and (2) who will consume the root messages.

A communication intention is defined as an intention of the speaker which is expected to be recognized during communication. It declares what information is to be expressed (Melinger & Levelt, 2004). Pragmatic theory posits that communicative intention of the message producer is the "why"—namely, the motive—of the message, which drives the receiver's response (Hellbernd & Sammler, 2016).

Human communication crucially depends on the existence of communicative intentions, which exist in the minds of message producers, and about which message receivers make inferences (Haugh, 2008). A message producer must have a communicative intention if her message is to count as communication (Recanati, 1986). For the communication to be effective, all partners involved in the conversation must agree on the communicative intention of the message (Ebert et al., 2018; Žegarac & Clark, 1999). Misunderstanding of the communicative intention can lead to misinterpretations (Ebert et al., 2018; Haugh, 2012).

In modern pragmatics, Grice is one of the pioneers to conceptualize communicative intentions. In his seminal work (Grice, 1957), Grice argued that a speaker *S* meant something by *x* if and only if *S* "intended the utterance of *x* to produce some effect in an audience by means of the recognition of this intention'' (Grisce, 1957, p. 385). The Gricean communicative intention is an intention such that (a) a message producer typically (but not necessarily) has this intention, and (b) a message counts as a part of communication if and only if it makes the intention manifest (Recanati, 1986).

The Gricean conceptualization above emphasizes receivers' recognition of communicative intentions, which differentiate communicative intentions from psychological states of individuals such as emotions and attitudes (Hellbernd & Sammler, 2016). Communicative intentions might be driven by message producers' emotions and attitudes. However, the former is considered as the goals of language that are expected to be recognized by the receivers and to influence their communicative reactions. In comparison, basic emotions and attitudes do not necessarily need a partner to be displayed or influenced (Hellbernd & Sammler, 2016; Wichmann, 2000).

3.2.1.2 Speech Acts of Producers

In pragmatics, it is not words, but the words' linguistic context that conveys the communicative intentions of messages (Searle & Searle, 1969). Such context is the speech acts of producers. Speech acts, also called language acts or linguistic acts (Searle, 1965), are defined as linguistic acts performed in the process of speaking (Ilyas & Khushi, 2012; Sadock, 2004), which can be used to conceptualize speech as action (Vosoughi & Roy, 2016). Examples of speech acts include: to express our emotions, start an argument, insult someone (Ilyas & Khushi, 2012), ask for information, agree with the partner, state facts, and express opinions (Albright et al., 2004; Ebert et al., 2018).

Speech acts serve to convey the communicative intention of a message in a discourse (Searle & Searle, 1969; Villarroel Ordenes et al., 2017; R. Zhang et al., 2011). People construct their conversations through a set of speech acts to yield a specific communicative intention (Ebert et al., 2018). Performing a speech act is essentially expressing a certain intention in uttering certain words. A message producer's communicative intention determines what speech acts she would take to perform (Sbisà, 2001). In Gricean conceptualization, communicative intentions must be recognized by receivers. From speech acts' point of view, an intention is recognizable if the producer chooses the certain speech acts to convey it (Bach, n.d.).

Speech acts are units of language discourse that provide both meaning and reality (Carr et al., 2012) and lie central to communication (Grundlingh, 2018). Making a speech means to make speech acts (La Rocca, 2020; Searle & Searle, 1969). Speech acts help the speaker or writer to communicate a message and enable the listener or reader to understand the message being received (Grundlingh, 2018). By making different types of speech acts, an individual is enabled not only operate within the world but also interact with the world around. In doing so, the

individual affects the attitudes and actions of those with whom he or she interacts (Carr et al., 2012; Fitch & Sanders, 2004).

3.2.1.3 Speech Act Theory

Speech acts were first defined and explained in the Speech Act Theory (SAT) (Austin, 1975; Chandler et al., 2018), one of the most influential linguistic theories to study language-in-use (Ludwig & de Ruyter, 2016). SAT is built upon the main assumption that saying is equivalent to doing (La Rocca, 2020) and that utterances are actions (Chandler et al., 2018; Schegloff, 1988). It conceptualizes all forms of speech as acts (Ludwig & de Ruyter, 2016) and asserts that words perform actions, and, in this way, language can be performative (Austin, 1975; Chandler et al., 2018). In this context, a speech act reflects both the utterance of language and the influence of the utterance on audience (Bach, n.d.; Chandler et al., 2018).

SAT addresses how word categories and sentence constructions used in people's everyday language give insights into their communicative intentions, perceptions and identities (Bagozzi et al., 2007; Ludwig & de Ruyter, 2016). The central premise of SAT is that language construction in speech or writing—through words, sentences, and interactional message exchanges—conveys a message producer's underlying meaning and intention (Austin, 1975; Ludwig & de Ruyter, 2016; Searle, 1976). It suggests interpretation of communicated words require recognition of a higher-order linguistic context (Ludwig & de Ruyter, 2016).

SAT suggests that there are three levels of speech acts (Austin, 1975; Evans, 2016; Grundlingh, 2018): the locutionary acts, illocutionary acts, and the perlocutionary acts. Locutionary acts refer to the production of sounds and words. Illocutionary acts refer to performing one of the functions of language, such as the act of saying or writing something, and expressing the intended meaning innate to communication (Ludwig & de Ruyter, 2016; Sbisà, 2001). Perlocutionary acts refer to the effects (intended and unintended) that result from the other two levels of speech acts (Grundlingh, 2018; Ludwig & de Ruyter, 2016). Essay II follows prior work (e.g., Carr et al. 2012; Ilyas and Khushi 2012; Searle 1965, 1976) to focus on illocutionary acts; the speech acts used in this study in relation to producers' communicative intention belong to illocutionary acts.

3.2.1.4 Measuring Producers' Communicative Intentions using Speech Acts

In Essay II, producers' communicative intentions are measured based on the speech acts they use in the root messages. Overall, communicative intention can be considered a type of

behavioral intention. A behavioral intention is an anticipated behavior. It represents someone's expectancies about some behavior under certain circumstances; it can be operationalized as the likelihood to perform specific actions (Fishbein & Ajzen, 1977; Lam & Hsu, 2006). A behavioral intention reflects the degree to which a person has formulated conscious plans to perform or not perform some specified behavior (Warshaw & Davis, 1985). In the case of communicative intention, the behavior or act is the speech act that the producer uses to convey the communicative intention. Literature suggests that individuals' behavioral intentions can be measured using various methods (e.g., Lam and Hsu 2006; Luarn and Lin 2005). In particular, messages users created on social media have been used as a medium to infer their behavioral intentions, such as their voting intentions (Lampos et al., 2013), purchase intentions (Ding et al., 2015), or any intentions (Z. Chen et al., 2013). In the case of communicative intentions, analysis of speech acts has been found useful for improving understanding of communicative intentions (Abbas et al., 2018; Te'eni, 2006). In particular, producers' communicative intentions are conveyed by the speech acts they choose (Searle & Searle, 1969; Villarroel Ordenes et al., 2017; R. Zhang et al., 2011). In turn, speech acts are important for understanding producers' communicative intentions (Abbas et al., 2018; Te'eni, 2006).

In this essay, producers' communicative intentions are measured based on Searle's taxonomy of speech acts (Nastri et al., 2006; Searle, 1976), which has been applied widely in social media research (e.g., Carr et al. 2012; Ilyas and Khushi 2012; Vosoughi and Roy 2016). In this taxonomy, five classes of speech acts are recognized—assertive acts, directive acts, commissive acts, expressive acts, and declarative acts (definitions and examples speech acts of these classes are presented in Table 16). Literature has suggested that these categories reveal the intentions of speakers (Nastri et al., 2006; Te'eni, 2006). On this basis, each category in the taxonomy transform to a communicative intention. For instance, the class assertive acts refer to statements of certain fact, trying to make the receiver to form or attend to a belief. This class is transformed to the intention to get the receiver to believe. The class directive acts refer to speech acts the sender uses to get the receiver to perform some activities (i.e., a command). This class is transformed to the intention to get the receiver to do something. The intentions from other classes are shown in Table 16.

Very few examples of taxonomies of communicative intentions can be found in literature (e.g., Coggins and Carpenter 1981; Ebert et al. 2018). The proposed set of intentions are in a

similar style like those. However, the existing taxonomies are found either published without an applicable measure (e.g., Ebert et al. 2018 does not provide any measure for its intention set), or only applicable to a special group of people (e.g., intention set by Coggins and Carpenter (1981) was designed for measuring children's intentions).

3.2.1.5 Producers' Communicative Intentions and Misinformation

As discussed above, producers' communicative intentions are represented using the classes of speech acts in Searle's taxonomy. By design, these classes are distinct in three dimensions (Searle, 1976). First of all, speech acts in different classes carry different purposes (called illocutionary points in the original work). For example, the purpose of an assertive speech act is representing (true or false, accurate or inaccurate) how something is. Differently, the purpose of a commissive speech act is to represent an undertaking of an obligation by the producer to do something. Secondly, some classes of speech acts are different in the direction of fit between words and the world. In other words, some categories express an intention to get the message content to match the world, while others to get the world to match the message content. For instance, assertive speech acts are in the former category, whereas commissive and directive speech acts are in the latter.

Third, different classes of speech acts express different psychological states of message producers. For example, a message producer who uses assertive speech acts to, e.g., state, explain, assert, or claim specific belief *p* expresses that she believes *p*. A producer who uses directive speech acts to, e.g., order, command, or request receiver *r* to do *a* expresses her desire (want or wish) that *r* does *a*. In addition, different speech acts—regardless of in which category they fit in—may present the same communicative intention with differing strength. "I suggest we go to the movies" and "I insist that we go to the movies", or "I guess Bill stole the money" and "I solemnly swear that Bill stole the money" present the same intention with varying degree of strength or commitment (Searle, 1976, p. 5).

On the pragmatic level, messages are composed of speech acts. Given that speech acts in different classes may be produced under different psychological states, present different directions of fit, and carry different types and strength of intentions, the resultant messages can also be different in the same dimensions. Differences in these dimensions can further impact how broadly the messages would propagate and who would like to pass on the messages. For instance, the neural-psychological states of people can impact their willingness in spreading

certain messages (e.g., Alhidari et al. 2015; Falk et al. 2012, 2013; McNeill and Briggs 2014). Also, users' communicative intentions also influence the likelihood they participate in message spreading (Alhidari et al., 2015; Yu Liu et al., 2016; H. Zhao et al., 2014).

3.2.2 Production Approaches of Misinformation

3.2.2.1 Approaches of Misinformation Production

There are three different views to understand the approaches to misinformation production on social media—to view the approaches as social media activities, as information manipulation, or as interpersonal deception. Existing research dominantly views the production approaches as social media activities conducted by users (namely, producers) intentionally or unintentionally. Producers try to exploit all the social media resources available to them in order to facilitate the spread of misinformation. In this view, misinformation can be created by producers based on sensational events or public issues already well-known on social media (Ahmed et al., 2020; Allcott & Gentzkow, 2017; Chiou & Tucker, 2018). During the production of misinformation, producers can amplify or frame (i.e. delimit the scope of) the events or issues strategically by distorting information from mainstream media, in order to interpret the situation in favor of their purposes (Marwick & Lewis, 2017). Variants of the same story can be created and delivered to social media in the same timeframe in order to create social memes, which can be passed along across social networks, growing into social phenomena (Qazvinian et al., 2011). To maximize the spread of their stories, multiple producers can compose their stories online collaboratively (Jenkins, 2009). It can be seen that this view of production approaches is based on lots of information not available in the content of the root messages. This would create difficulties in measuring the approaches in research. Also, this view is neither systematic, nor supported by any mature theories, which creates difficulties in relating to the production approaches to the socio-psychological intentions of producers.

Alternatively, the production approaches can be viewed as strategies of information manipulation. In the context of political propaganda, information manipulation is defined as "the intentional and massive dissemination of false or biased news for hostile political purposes" (Vilmer et al., 2018, p. 12). More generally, it refers to any activity, either interpersonal or public, of creating and emitting false information based on fact (Edmond, 2013; Goldman & Slezak, 2006; McCornack, 1992). The Information Manipulation Theory (IMT) (McCornack,

1992) describes the linguistic methods of information manipulation. It can be used to represent the approaches to producing misinformation on social media.

IMT describes a multidimensional approach (Yeung et al., 1999) to producing deceptive messages. The theory integrates Grice's (Grice, 1989) theory of conversational implicature with analysis of deception as information control (e.g., Bavelas et al. 1990; Bowers et al. 1977; Metts 1989; Turner et al. 1975; Yeung et al. 1999). In particular, IMT rests on Grice's (Grice, 1989) Cooperation Principle and its maxims as a foundation for describing a variety of forms of manipulation activities. IMT considers manipulation of information as arising from covert violations of one or more of Grice's four maxims (quality, quantity, relevance, and manner): Covert violations of quality leads to the falsification of information. Covert violations of quantity can introduce lies of omission. Covert violations of relevance involve manipulation by evasion. Covert violations of manner result in deception by equivocation.

Finally, production approaches can also be viewed as strategies of interpersonal deception. Deception is described as a deliberate act performed by a sender to create beliefs in the receiver; such beliefs are contrary to what the sender believes, in order to put the receiver at a disadvantage (Burgoon & Buller, 1994; Wise & Rodriguez, 2013). Essentially, deception is an attempt for a deceiver to mislead the deceivee (Wise & Rodriguez, 2013). During deception, the deceiver creates a facade of truth in the deceivee's mind by sending false information, neglecting truthful information, or changing the surrounding environment, in order to establish a deceptive framework (Buller et al., 1994).

The Interpersonal Deception Theory (IDT) (Buller & Burgoon, 1996) provides a powerful lens to inspect the essential aspects of interpersonal communication and deception (Kirk, 2015). It was proposed in order to account for credible and noncredible communication in interpersonal contexts, approaching "the issue relationally, considering deceptive interchanges from a dyadic and dialogic rather than monadic and monologic perspective" (Buller & Burgoon, 1996, p. 204). IDT suggests that deception can succeed in two cases (Ojebode, 2012): First, respondents' persistent expectation for others to tell the truth can make them believe the deceptive messages while ignoring possible cues of deception. Second, skilled deceivers can perform specific strategies, verbal or non-verbal, which make their messages appear believable. Deception often fails due to deceivers' intentions. Deceivers driven by selfish intentions fail more easily than deceivers with altruistic intentions, for selfish intentions are harder to hide

(Littlejohn & Foss, 2010). Furthermore, IDT specifies six semantic-syntactic approaches that can be used to create misleading information for the purpose of deception (Burgoon et al., 1996) (presented in Table 5): quality (truthfulness) manipulations, quantity (completeness) manipulations, clarity (vagueness and uncertainty) manipulations, relevance manipulations, depersonalism (disassociation) manipulations, and image- and relationship-protecting behavior. These approaches can be recognized through, and thus measured from, the linguistic characteristics of the produced messages (Burgoon et al., 1996; L. Zhou et al., 2004).

3.2.2.2 The Interpersonal Deception View

In this study, production of misinformation is viewed as interpersonal deception, and the approaches to misinformation production are represented using the semantic-syntactic approaches of deception recognized by IDT. The deception view supported by IDT can help to relate Production Approaches to the other constructs in downstream—Root Message Propagation and Consumer Profile. IDT can map production approaches to their linguistic cues reflected in root messages. Meanwhile, it has been shown that linguistic characteristics of misinformation messages have indication on how the messages propagate and be consumed by receivers (Shu, Wang, et al., 2017).

Indeed, IDT was originally designed for modeling deception, which is typically conducted by people with deceptive intention, whereas a part of misinformation on social media (i.e., the non-disinformation) is not produced under deceptive intention. However, it has been suggested that the deceptive approaches described by IDT, including varying the amount of information, veracity of information, relevance of information, and clarity of information, are also applicable to producing any type of misinformation—intentional or unintentional (L. Zhou & Zhang, 2004). On the other hand, deception activities can also be conducted without the intention to cheat or lie. For example, there exists unintentional deception, which might result from misunderstandings from either parity of deception (Banja, 2010; Penco & Beaney, 2009; Shim & Arkin, 2013; Tilwick, 1975; Ward & Hexmoor, 2003). Furthermore, IDT has been applied in modeling and analyzing general misinformation (Kirk, 2015; L. Zhou & Zhang, 2007). Based on these considerations, I apply the IDT-supported strategies on general types of misinformation in this study.

3.2.3 Opinion Polarization on Social Media and Its Measures

Originally, polarization, or opinion polarization, refers to a phenomenon that a group of people are deeply divided on specific issues and the divisions about such issues tend to be more and more aligned with these people's identities, such as their partisan identities (Bail et al., 2018; Spohr, 2017). (Degree of) polarization is "the extent to which opinions on an issue are opposed in relation to some theoretical maximum" (DiMaggio et al., 1996, p. 693). In social networks, polarization is described as a status that users in different communities within a social network present distinct, even opposite opinions on a common topic; different communities tend to hold negative attitudes towards each other (Conover et al., 2011). It happens when people with similar opinions strengthen the belief of each other. For example, people who are opposed to the minimum wage might become even more opposed after talking to each other; people who support gun control becomes even more supportive after seeing other people's supportive attitudes (Sunstein, 1999). Depending on the topics, there are different dimensions of polarization, such as political polarization, ideological polarization, and issue-based polarization (Lee et al., 2014). Generally, a higher degree of polarization will be asserted if (1) user opinions in the same community are closer to each other, and (2) user options in different communities are more distant from each other (Conover et al., 2011; Matakos et al., 2017; Morales et al., 2015).

Polarization on social media has been identified as a pressing social problem ever since the main social media platforms such as Twitter and Facebook gained mass popularity at the beginning of 2010s. The fundamental mechanism leading to polarization is that posts (e.g., replies, comments) between like-minded users on social media can strengthen group identity, whereas posts between different-minded individuals can reinforce in-group and weaken out-group affiliation (Yardi & Boyd, 2010). This polarizing mechanism is even more powerful with social networks formed by pure re-posting (e.g. retweeting on Twitter and forwarding on Facebook). For instance, politically polarized networks of retweets usually present a highly segregated partisan structure with extremely limited connectivity between left- and right-leaning users (Conover et al., 2011). Users tend to dislike, even loathe, out-group users, showing biased views of their opponents (Iyengar et al., 2012). Their willingness or tendency to shift their allegiance to their opinion and groups is minimal (Gruzd & Roy, 2014). Over the years, polarization on social media has become increasingly intense in multiple dimensions, such as the way people follow political and media accounts, whether they re-post content from the other side

of opinion, and the hashtags they use etc. (Garimella & Weber, 2017). Despite its spreading social effect, polarization is unevenly prevalent among opinion groups. For example, the distribution of polarization-indicating content (e.g., fake news, sensationalist messages, conspiracies) is unevenly spread across the ideological spectrum (Narayanan et al., 2018).

More and more evidence has been found, which renders polarization on social media as a predictor of the prevalence of misinformation (Allcott & Gentzkow, 2017; Bessi et al., 2015; Tucker et al., 2018; Vicario et al., 2019). On the one hand, it has been observed that users in polarized networks are more prone to diffuse false information (Bessi et al., 2015). On the other hand, polarized users tend to believe misinformation because their vulnerability to false information is increased by directionally motivated reasoning, which occurs when their belief is biased due to polarization (Tucker et al., 2018). In turn, opinion homophily and polarization are considered key metrics to recognize communities where false or misleading information is likely to spread (Bessi et al., 2015). Users' degree of polarization has also been used to determine in advance potential topics of fake news (Vicario et al., 2019).

There are two mechanisms that contribute to the growth of polarization of social media: echo chamber and filter bubble. Echo chambers occur when external information sources, such as the political opinion leaders and media outlets, may feed slanted messages, such as partisan propaganda, to the audience on social media. This can cause ideologically one-sided information exposure confined to a small, but highly involved and influential, segment of social media users (Prior, 2013). Due to the effect of confirmation bias (Knobloch-Westerwick & Kleinman, 2012), these users tend to listen to information that supports and adheres to their beliefs, and to form communities sharing the same view—such communities are called echo chambers (Bessi et al., 2016). Inside an echo chamber, users' emotional behavior is strengthened by their involvement, and more involved users are found to show a faster shift towards the negative emotion (Vicario et al., 2019). Strong emotion gradually leads to polarization of users in opponent communities.

Concurrently to the forming of echo chambers, platform-embedded algorithmic curation and personalization systems smartly filter out content that users dislike while injecting more and more messages users like to open and view, placing users in a filter bubble of content. Such mechanism further decreases users' chance of encountering ideologically cross-cutting content (Spohr, 2017). Altogether, echo chambers and filter bubbles jointly foster users' negative

attitudes towards those with different opinion and their confirmative attitudes towards the like-minded, creating the more and more polarized social media (Beam et al., 2018).

There are two types of measures for evaluating the degree of polarization of a group of users: structural measures and content-based measures. Both types of measures quantify polarization of users by inspecting if they can form clear, cohesive communities within the same social network. The structural measures inspect the network structures of the communities and assert a high degree of polarization if these communities are closely connected internally, but loosely connected to each other (Conover et al., 2011; Guerra et al., 2013). Differently, the content-based measures inspect the contents created or listened to by each community in the group. A high degree of polarization will be asserted if users in the same community are engaged in contents with the same opinion, while users in different communities are engaged in contents with opposite opinions. Opinions of contents can be measured by their sentiment, emotion, topics, and hashtags used (Barberá, 2014; Garimella & Weber, 2017). Several studies go further to design integrated polarization measures that incorporate the structural and content-based solutions. For instance, the polarization index (Morales et al., 2015) quantifies the polarization observed in a Twitter social network given the network structure and the opinions of the users in the network. This index is a function of the opinion of each user, which further depends on the opinion of the opinion leader the is connected to and how close this connection is. Opinion of an opinion leader is determined by the frequency of keywords he or she used in past tweets. Another design of polarization index (Matakos et al., 2017) refines the earlier design by considering the forming of user opinion. Each user has internal and expressed opinions. The latter is an aggregation of the former and the expressed opinion of each connected peer. The polarity of the entire network depends on the expressed opinions of all the users.

The above research efforts were made to address polarization of groups of people. In parallel, research interest has also been directed to polarization of individuals (Borella & Rossinelli, 2017; Boxell, 2020; Boxell et al., 2017; Del Vicario, Bessi, et al., 2016; Marozzo & Bessi, 2018). This body of research aims to define the polarization or polarity of an individual's opinion and quantify it based on the opinion footprints of the individual. Such footprints can be answers made by individuals in ideological surveys (Borella & Rossinelli, 2017; Boxell, 2020), or likes they made (Del Vicario, Bessi, et al., 2016) or sentiments they expressed (Marozzo &

Bessi, 2018) on social media. Mathematically, individual polarization is typically a function of the individual's opinion measured in such footprints.

3.2.4 Measures of Opinion Polarization

Existing measures of individual polarization have largely ignored the status of groups. These measures are almost purely computed based on the strength of opinions (measured by e.g., user's use of sentiment words or online behavior) of the individuals, without considering the opinion of the group to which the individuals belong. In turn, those measures are essentially not different from measures of personal sentiment or attitudes (e.g., Hutto and Gilbert 2014; Thelwall 2017; Thelwall et al. 2010).

However, in one of the seminal studies in this field, opinion polarization is defined as the extent to which people's opinions are opposed with respect to a certain maximum (DiMaggio et al., 1996), which suggests that polarization should be considered based on the relative position of opinions of multiple people, not the absolute strength of opinion of any individual alone. In addition, opinion polarization of a group have influence on opinions of individuals in the group (Beam et al., 2018; Del Vicario, Vivaldo, et al., 2016; Spohr, 2017). Therefore, I propose that individual opinion polarization should be redefined considering (1) opinion of the individual, as in the existing measures of individual opinion polarization, and (2) degree of opinion polarization of the group, as well as the opinion extremes existing in the group, and (3) distance of the individual opinion relative to these group extremes.

I propose the construct Producer Opinion Polarity. Opinion polarization has been found to be able to manifest in the content of messages generated by users, reflected in the emotions, sentiments, and topics expressed by the messages. In the emotion and sentiment dimension, when users discussing certain topics—especially controversial political topics—become more and more polarized, the emotions and sentiments they express can become increasingly negative (Sobkowicz & Sobkowicz, 2012). However, users in different camps in a polarized environment may present differential tendency. For instance, politically liberal-leaning users tend to elicit more positive collective emotions than conservative-learning users (Garcia et al., 2012). During the forming of the polarized environment, users with strongest individual polarity (e.g., some politicians) are usually the key players: in the initial phase of polarization, they carry strong emotions and pass both their opinions and emotions to other users in the network who have low emotions (Alamsyah & Adityawarman, 2017). This triggers the development of echo chambers,

in which users become more and more polarized, engaged, and (negatively) emotional (Del Vicario, Vivaldo, et al., 2016).

In the topic dimension, opinion polarization can be reflected in the topics users choose to discuss in their messages on social media. Opinion polarization is changing many aspects of people's life. It affects the social relations we seek to enter into, such as our friendships, romantic relationships, or marriages (Iyengar et al., 2019; Nicholson et al., 2016). For example, marriages are found to be more politically homogeneous than by chance (Stoker & Jennings, 1995). Also, polarization distorts the economic behavior of people (Iyengar et al., 2019). For instance, some taxi drivers accept lower prices from co-partisans and demand higher prices from counter-partisans (Michelitch, 2015), while employers tend to accept resumes from co-partisans rather than from counter-partisans (Gift & Gift, 2015). With so many aspects of life being changed, people might start to develop new lifestyles, life philosophies, and states of mind. Naturally, people would talk about different topics on social media as they might be interested in matters they did not care about before.

Furthermore, literature suggests that characteristics of message content can impact propagation of the messages. Content features—linguistic and semantic—are among the standard features used to predict message propagation (Shu, Wang, et al., 2017). For misinformation messages, unique characteristics have been found in their content and propagation patterns, which distinguish them from true information. These content characteristics include, among others, the use of sentiment words (e.g., positive and negative emotion words), cognitive words (e.g., cause, know, ought), and tentative words (e.g., may be, perhaps, guess). The propagation characteristics include the sizes, shapes, density, and clustering of the propagation cascades (Kwon et al., 2013).

In addition, many features of the message content have been found to predict the propagation of messages directly. These features include many lower-level metrics, such as number of hashtags, mentions, URLs, trending words, length of the tweet, is the tweet a reply, and the actual words measured in TF-IDF (Hong et al., 2011; Kupavskii et al., 2012; Petrovic et al., 2011). On the semantic level, emotions expressed in messages, such as disgust, fear, and anger, are found to be able to increase the message endorsement during propagation (Hochreiter & Waldhauser, 2014). Sentiment of messages represented by the use of positive/negative terms and smileys can influence the size of propagation cascades of messages (Kupavskii et al., 2012).

Topics involved in messages can influence the growth of cascade size in the near future (Hong et al., 2011).

Lastly, Consumer Profile in this study mainly comprises four types of user features—their emotions, sentiments, demographics, and engagement on the platform. Extensive research has shown that users express different sentiments (Hutto & Gilbert, 2014; Neri et al., 2012; Nguyen et al., 2012; Yoo et al., 2018) and emotions (Brady et al., 2017; De Choudhury et al., 2013; C. Yang et al., 2009) depending on what they receive on social media. Meanwhile, different messages might be particularly interesting for users from specific demographic groups (Mellon & Prosser, 2017; Mislove et al., 2011; Sloan et al., 2013) and engagement levels (de Oliveira et al., 2016; Khan, 2017; Leung et al., 2013). Therefore, the content of misinformation root messages can influence the persona of consumers who consume these messages.

3.3 Hypotheses and Research Model

The research model of Essay II is shown in Figure 5. In summary, there are four main constructs:

Producer Communicative Intentions

Producer Communicative Intentions reflect the communication intentions of producers, namely what the producers intend to express through the root message, indicating what they will do, what they want the receivers to know, or how they want the receivers to react. Producers' communication intentions are seen as the pragmatic forces that drive misinformation diffusion. They are measured using the five classes of speech acts in Searle's taxonomy.

The taxonomy was designed in a way that different classes of speech acts have distinction in aspects of the linguistic purposes they express, their directions of fit between words and the world, and the expressed psychological states of message producers (Searle 1976). Thus, root messages composed of different classes of speech acts (in other words, carrying different types of communicative intentions) also have differences on the aspects. Such differences can further impact how broadly the messages would propagate and who would like to pass on the messages (e.g., Alhidari et al. 2015; Falk et al. 2012, 2013; Liu et al. 2016; McNeill and Briggs 2014; Zhao et al. 2014).

As described in Table 16, there are five classes of communicative intentions or speech acts based on the Speech Act Theory, including: representative (REP), directive (DIR), commissive (COM), expressive (EXP), and declarative (DEC). In addition, literature measuring

speech acts on social media content (Abbas et al., 2018; Te'eni, 2006) suggests adding another class, quotation (QUO), meaning that the text intends to reflect original messages from other speakers. To measure Producer Communicative Intention, I followed prior studies (e.g., Carr et al. 2012; Chandler et al. 2018; Ilyas and Khushi 2012; Vosoughi et al. 2018; Zhang et al. 2012) to annotate the occurrence of these six types of intentions in the root messages I collected. Each root message was then associated with a binary vector of six digits, with each digit representing the occurrence of a type of intention. Finally, DEC was removed from the data because it only appeared a few times. In addition, the number of intentions detected was also added as a variable associated to Producer Communicative Intention.

Table 16. Measures of Producer Communicative Intentions

Variables	**Communicative Intention**	**Properties of Corresponding Speech Act** (Carr et al., 2012; H. Schiffman, 1997; Searle, 1976)	**Properties of Communicative Intention**
REP (binary)	Representative intention	Statements of fact, getting the viewer to form or attend a belief	To make the receiver believe some statements of fact
DIR	Directive intention	The sender uses this to get the receiver to do something (i.e., a command)	To make the receiver do something
COM	Commissive intention	The sender commits himself to do something	To make the producer herself do something
EXP	Expressive intention	Sender expresses feeling toward (though not necessarily about) the receiver	To let the receiver recognize the feeling expressed by the producer
QOU	Quotation intention	Sender wants to share an original piece of statement reportedly from some other source	To let the receiver believe the statement and the reported authorship of the statement
Count (int)			Number of intentions held simultaneously

Production Approaches

Production Approaches refers to the semantic and syntactic approaches taken by producers to producing misinformation. It reflects the semantic-syntactic forces that drive misinformation production and diffusion. It is measured using the six semantic-syntactic approaches to interpersonal deception based on IDT (Zhou et al. 2004). IDT indicates that different production approaches have unique linguistic cues reflected in root messages. Meanwhile, it has been shown that linguistic characteristics of misinformation messages have indication on how the messages propagate and be consumed by receivers (Shu, Wang, et al., 2017).

The six types of approaches are defined in Table 17. They are qualitative manipulation (QUA), quantitative manipulation (QUT), relevance manipulation (REL), clarity manipulation (CLA), depersonalization (DEP), and image protection (IMG). Similarly, I annotated these approaches in the root messages and assigned and binary vectors to represent them. Finally, CLA was removed from the data because it only occurred a few times. In addition, the number of approaches detected was also added as a variable under Production Approach.

Table 17. Measures of Misinformation Production Approaches

Variable	**Production Approach**	**Linguistic Cues** (L. Zhou et al., 2004)
QUA (binary)	Quality (truthfulness) manipulations	No clear cues, but can be measured by comparing to the true information given in the corresponding posts on fact checkers.
QUT	Quantity (completeness) manipulations	Fewer words and sentences; root messages may be incomplete syntactically, by giving perceptually less information, or semantically, by failing to present actual detailed content such as factual statements.
REL	Relevance manipulations	Root messages are semantically indirect (e.g., making polite speech) or irrelevant (e.g., providing irrelevant details), or syntactically indirect (e.g., following a question with a question)

DEP	Depersonalism (disassociation) manipulations	Using nonimmediate language such fewer first person pronouns, more passive voice, more second person pronouns
IMG	Image- and relationship-protecting behavior	Avoidance of discrediting information (e.g., admitted lack of memory, expressions of doubt) and avoidance of negative affect in one's language (partially intended to cover any accidental betrayal of true feelings of guilt, fear of detection, etc.)
Count (int)		Number of approaches used simultaneously

Producer Opinion Polarity

Producer Opinion Polarity reflects the polarity of producers' opinion on the COVID-19 pandemic. Literature review in previous subsections suggests that opinion polarity or polarization depends on two factors: one's attitude towards a topic, and one's strength of sentiment and emotion expressed. I extend ideas in prior work (Hutto and Gilbert 2014; Thelwall 2017; Thelwalls et al. 2010) to measure Producer Opinion Polarity with two numeric variables, Root Message Polarity and Producer Polarity.

Literature suggests that our sentiment, emotion, and opinion over a specific subject expressed in text (e.g., in tweets and Facebook posts) are closely interrelated (B. Liu, 2012). In particular, social media users' sentiment and emotion expressed in the content they produce are important representatives of their opinion (B. Liu, 2012). Their sentiment and emotion can predict the polarity of their opinion (Borella & Rossinelli, 2017; T. Chen et al., 2019). Following the prior research like these, I compute both measures of Producer Opinion Polarity—the Root Message Polarity and Producer Polarity—using producers' sentiment and emotion expressed in the root messages they produce and the user description they compose, respectively.

Root Message Polarity is designed to reflect producers' opinion on COVID-19 expressed in their root messages. To compute this measure, I first annotated the common, COVID-related topics talked about by the root messages I collected (n=927). In result, there were around 30 different topics detected from the root messages, with each root message talking about one or

more of them. Then, I separated the topics into two groups. One group of the topics reflect that the authors have generally *active* attitudes towards the pandemic. Topics in the active group are such as pro-vaccine, pro-mask, admitting the severeness of the pandemic situation, or admitting the danger of this disease. Another group of the topics reflect an *inactive* attitude towards COVID, such as anti-vaccine, anti-mask, believing that it is not worse than a flu, or believing that the pandemic is a hoax. Then, the root messages talking about COVID-active topics were labeled COVID-active root messages, while those talking about COVID-inactive topics were labeled COVID-inactive root messages. Next, the absolute values of the sentiment scores and emotion scores of each root message were extracted and averaged into a single number in [0, 1]. For the COVID-active root messages, their Root Message Polarity equals to the aggregate (i.e., averaged) sentiment/emotion score, a positive number in [0, 1]. For the COVID-inactive root messages, their Root Message Polarity equals to this score multiplying -1, namely, a negative number in [-1, 0]. It can be seen that Root Message Polarity reflects the opinion and strength of sentiment and emotion expressed in root messages simultaneously.

In the computation of the Root Message Polarity, the absolute values of the sentiment and emotion scores of the root messages are taken, without considering the positivity and negativity of the root messages' sentiment and emotion scores. This is because the Root Message Polarity aims to measure root messages' opinion on the given subject (i.e., how active or inactive they are towards the fight against COVID-19). A root message can express the same opinion by using positive OR negative sentiments/emotions. The positivity and negativity of sentiment/emotion of a root message does not necessarily influence the opinion expressed by this root message. What really matters is the strength of its sentiment/emotion, namely, the absolute values of the sentiment and emotion scores.

On the other hand, Producer Polarity was computed by averaging the sentiment and emotion scores of each producer's user description. For scores of negative sentiment or emotions, such as anger, disgust, and fear, they were given a negative sign before averaged with the other scores. Consequently, if the value of Producer Polarity is negative, that means the producer expressed more negative sentiment or emotion then positive ones in their user description. Overall, Producer Polarity is a number in [-1, 1] that measures a producer's sentiment and emotion expressed in the user description.

In the computation of the Producer Polarity, the signed sentiment and emotion scores of the user description are employed, which means that the positivity and negativity of the sentiment and emotion scores of producers' user description are considered. This is because the Producer Polarity aims to measure producers' sentiment and emotion in general. In this case, the positivity and negativity of the involved sentiment/emotion scores reflect the positivity and negativity of producers' sentiment/emotion in general.

With all the constructs and measures defined, I propose the following hypotheses, which are related to the research model depicted in Figure 5.

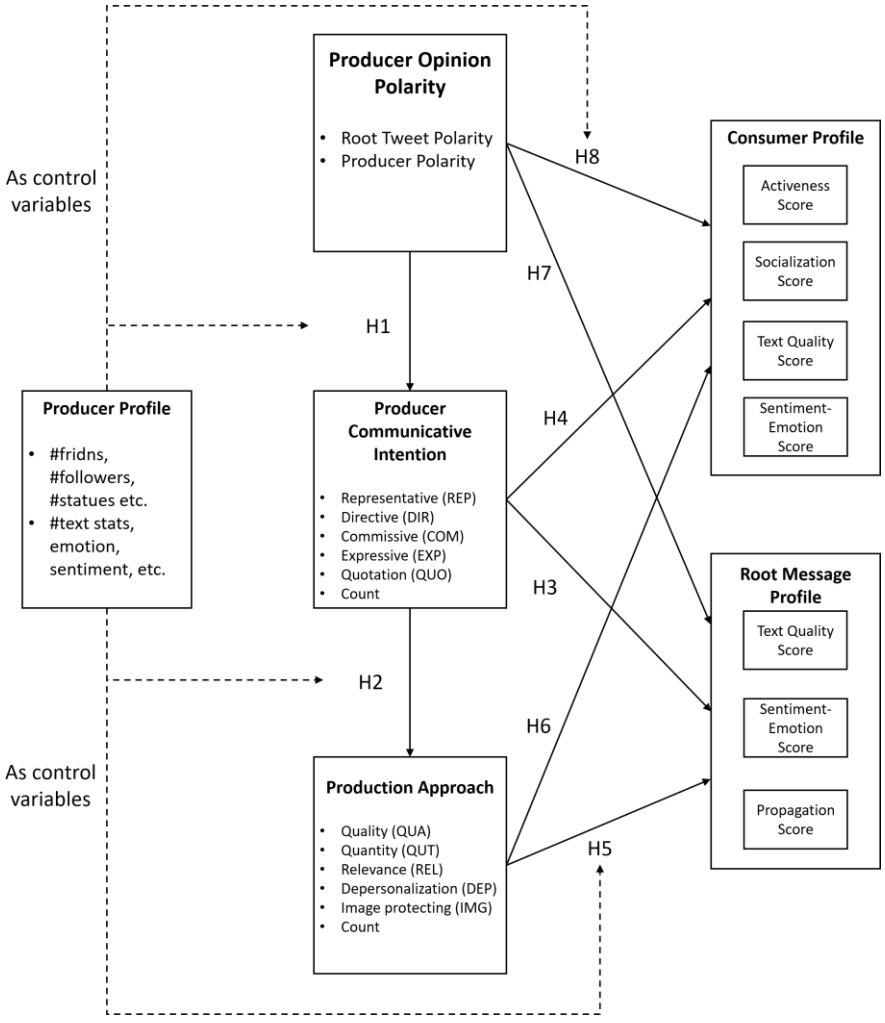

Figure 5. Research Model of Essay II

First, I posit that producers' opinion polarity is the driving force the ignites their intention of producing misinformation. As a result, producers' communication intention of expressing the misinformation on social media emerge. Therefore, I propose H1:

H1: Producers' opinion polarity influences their communicative intention to produce misinformation.

Then, I posit that producers' communicative intention is one of the factors that determines how they want to formulate the information they want to convey, namely, influencing which approaches they take to produce misinformation. Meanwhile, characteristics and the producers (i.e., their online profiles) and their opinion polarity might also impact their choice of production approach. Therefore, these two constructs need to be accounted for (as control variables). Thus, I propose H2:

H2: Producers' communicative intention of producing misinformation influences the approaches they take to producing this misinformation.

Furthermore, producers' communicative intention and their choice of production approach influence the profiles of the root messages produced by these producers and which users consume these root messages. Meanwhile, producers' profile and their opinion polarity might also exert certain effects on root messages and consumer. Thus, producers' profile and their opinion polarity need to be accounted for (as control variables). Therefore, I hypothesize as follows.

H3: Communicative intentions held by producers when they produce root messages influence the (1) text quality, (2) sentiment and emotion, and (3) propagation of the root messages produced by these producers.

H4: Communicative intentions held by producers influence which users consume the root messages produced by these producers; these users are characterized/differentiated by their (1)

activeness on the platform, (2) their degree of socialization, (3) the text quality of their user description, and (4) sentiment and emotion expressed in their user description.

H5: Production approaches chosen by producers when they produce root messages influence the (1) text quality, (2) sentiment and emotion, and (3) propagation of the root messages produced by these producers.

H6: Production approaches chosen by producers influence which users consume the root messages produced by these producers; these users are characterized/differentiated by their (1) activeness on the platform, (2) their degree of socialization, (3) the text quality of their user description, and (4) sentiment and emotion expressed in their user description.

Finally, producers' opinion polarity can directly influence the profiles of the root messages produced by these producers and which users consume these root messages. Meanwhile, producers' profile might also have effects. Thus, producers' profile needs to be accounted for (as control variables). Therefore, I hypothesize as follows.

H7: Opinion polarity held by producers when they produce root messages influence the (1) text quality, (2) sentiment and emotion, and (3) propagation of the root messages produced by these producers.

H8: Opinion polarity held by producers influence which users consume the root messages produced by these producers; these users are characterized/differentiated by their (1) activeness on the platform, (2) their degree of socialization, (3) the text quality of their user description, and (4) sentiment and emotion expressed in their user description.

3.4 Methodology

To complete Essay II, an empirical study has been conducted using the Twitter data described in Essay I. Statistical analyses were conducted on the root messages with misinformation (n = 628) in order to verify or reject the hypotheses. A series of regression analysis addressing the hypotheses between the constructs. All statistical analyses were conducted at the level of the root messages. The multinomial logistic regression was employed,

as the dependent variable involved in each hypothesis is of a nominal variable, representing the cluster assignment of a root message. The multinomial logistic regression function in R was used to perform the regression.

3.5 Analyses and Results

3.5.1 Relationship Among Producer Opinion Polarity, Producer Communicative Intention, and Production Approach (H1 and H2)

First of all, the linear model in Equation (11) is established and fitted on the collection of misinformation root messages:

$$ProducerComInt_{clu} = ProducerOpinionPolarity + ProducerProfile \quad (11)$$

$ProducerComInt_{clu}$ represents the clustered form of the variables under the construct Producer Communicative Intention, namely, Producer Communicative Intention Clustered. *ProduerOpinionPolarity* includes all the individual variables of the construct Producer Opinion Polarity and their order-two interaction terms. *ProducerProfile* includes all the individual variables under the construct Producer Profile, which is the set of control variables.

Table 18. Description of Clusters of Producer Communicative Intention Clustered

Cluster Label	Cluster Size	Cluster Characteristics	Avg. REP	Avg. DIR	Avg. COM	Avg. EXP	Avg. QUO	Avg. Count
0	145	High in DIR, Medium in EXP	0.91	1	0.09	0.628	0.745	3.379
1	282	Low in all intentions	0.741	0.121	0.039	0.05	0.798	1.755
2	201	High in COM, High in EXP	0.985	0	0.184	0.95	0.751	2.876

Table 18 describes the clusters computed for Producer Communicative Intention Clustered over the misinformation root messages. The variables of Producer Communicative Intention, namely, REP, DIR, COM, EXP, QUO, and Count, are considered as the six dimensions of each cluster. The clustering result by the K-Means algorithm (number of clusters set to 3) indicates that there are three main clusters (clusters 0, 1, and 2) taking up all the root messages of misinformation. According to the mean values of the six variables in each cluster, cluster 0 can be considered as representing those misinformation root messages (n = 145) that are highly likely to carry the communicative intentions of DIR and moderately likely to carry EXP. Cluster 1 represents the misinformation root messages (n = 282) that are less likely to carry any

communicative intention. Cluster 2 represents the misinformation root messages (n = 201) that are highly likely to carry the communicative intentions of COM and EXP.

Results of regression in Table 19 indicate significant influence of Producer Opinion Polarity upon Producer Communicative Intention (supporting H1). Both variables under Producer Opinion Polarity are found to be strongly associated with the cluster assignment of Producer Communicative Intention Clustered. Specifically, root message polarity (rooPol) is strongly, negatively associated with the odds of being assigned to clusters 1 (in relation to the odds of being assigned to cluster 0), and even more negatively associated with the odds of being assigned to cluster 2 (in relation to the odds for cluster 0). Meanwhile, producer polarity (prdPol) influences the odds of entering cluster 1 most positively, and the odds of entering cluster 2 moderately positively. Furthermore, the interaction term of these two variables strongly positively influences the odds for cluster 1 while strongly negatively influencing the odds for cluster 2.

Table 19. Regression of Producer Opinion Polarity on Producer Communicative Intention Clustered (H1)

IV's	Producer Opinion Polarity	
DV	Producer Communicative Intention Clustered	
Num. Obs.	591	
AIC	1267.666	
Cluster	#1	#2
Num. Sig. Main IV's	2	2
Num. Sig. Main IV's Inter.	1	1
Num. Sig. CV's	26	26
rooPol	-1.041+	-1.49+
prdPol	2.066+	0.664+
rooPol * prdPol	39.366+	-18.395+
p-value +: <0.05, ++: <0.01, +++: <0.001		

Then, the linear model in Equation (12) is established and fitted on the collection of misinformation root messages:

$$ProductionApproach_{clu} = ProducerComInt + ProducerProfile \quad (12)$$

ProductionApproach$_{clu}$ represents the clustered form of the variables under the construct Production Approach, namely, Production Approach Clustered. *ProducerComInt* includes all the individual variables of the construct Producer Communicative Intention and their order-two

interaction terms. *ProducerProfile* includes all the individual variables under the construct Producer Profile, which is the set of control variables.

Table 20. Description of Clusters of Production Approach Clustered

Cluster Label	**Cluster Size**	**Cluster Characteristics**	**Avg. QUA**	**Avg. QUT**	**Avg. REL**	**Avg. IMG**	**Avg. DEP**	**Avg. Count**
0	178	Low in approaches	0.652	0.489	0.039	0.596	0.045	1.82
1	310	High in QUT	0.839	1	0.171	0.89	0.594	3.51
2	140	High in QUA, High in IMG, High in DEP	1	0	0.179	0.964	0.907	3.064

Table 20 describes the clusters computed for Producer Approach Clustered over the misinformation root messages. The variables of Producer Approach, namely, QUA, QUT, REL, IMG, DEP, and Count, are considered as the six dimensions of each cluster. The clustering result by the K-Means algorithm (number of clusters set to 3) indicates that there are three main clusters (clusters 0, 1, and 2) taking up all the root messages of misinformation. According to the mean values of the six variables in each cluster, cluster 0 can be considered as representing those misinformation root messages (n = 178) that are less likely to be produced using any approach under Production Approach. Cluster 1 represents the misinformation root messages (n = 310) that are highly likely to be produced using the approach of QUT. Cluster 2 represents the misinformation root messages (n = 140) that are highly likely to be produced using the approaches of QUA, IMG, and DEP.

It can be seen from the results in Table 21 that variables of Producer Communicative Intention are significantly associated with Production Approach (supporting H2). Specifically, producers' likelihood of having a representative intention (high in prdInt_REP) or quoting intention (high in prdInt_QOU) when producing root messages is strongly, positively associated with the odds of them to enter cluster 2 of Production Approach Clustered. Their likelihood of having a directive (prdInt_DIR), commissive (prdInt_COM), or expressive intention (prdInt_EXP) is strongly, passively associated with their odds of being assigned to cluster 1 of Production Approach. If they have a high number of any intention (high in prdInt_count), their odds for cluster 0 is maximized. In addition, the interaction terms of any types of intentions can strongly influence producers' odds of going to a specific cluster. Most of the interaction terms positively influence the odds for cluster 2.

Table 21. Regression of Producer Communicative Intention on Production Approach Clustered (H2)

IV's	Producer Communicative Intention	
DV	Production Approach Clustered	
Num. Obs.	591	
AIC	1204.948	
Cluster	#1	#2
Num. Sig. Main IV's	6	6
Num. Sig. Main IV's Inter.	15	15
Num. Sig. CV's	28	27
prdInt_REP	56.681+	302.012+
prdInt_DIR	403.03+	213.439+
prdInt_COM	365.196+	-269.569+
prdInt_EXP	512.983+	435.893+
prdInt_QOU	22.294+	391.684+
prdInt_count	-456.547+	-295.514+
prdInt_REP * prdInt_DIR	4.776+	-32.525+
prdInt_REP * prdInt_COM	-31.307+	433.286+
prdInt_REP * prdInt_EXP	116.587+	196.085+
prdInt_REP * prdInt_QOU	535.637+	227.271+
prdInt_REP * prdInt_count	-57.606+	-49.724+
prdInt_DIR * prdInt_COM	-143.882+	-172.678+
prdInt_DIR * prdInt_EXP	4.79+	62.866+
prdInt_DIR * prdInt_QOU	-32.492+	53.602+
prdInt_DIR * prdInt_count	53.449+	82.204+
prdInt_COM * prdInt_EXP	-31.637+	55.748+
prdInt_COM * prdInt_QOU	-68.684+	43.611+
prdInt_COM * prdInt_count	89.687+	90.399+
prdInt_EXP * prdInt_QOU	79.263+	281.184+
prdInt_EXP * prdInt_count	-57.994+	-144.076+
prdInt_QOU * prdInt_count	-21.188+	-136.002+
p-value +: <0.05, ++: <0.01, +++: <0.001		

3.5.2 Influence of Producer Communicative Intention (H3 and H4)

First, the linear model in Equation (13) is established and fitted on the collection of misinformation root messages:

$$RootMessageProfile_{clu} = ProducerComInt + ProducerProfile + ProducerOpinionPolarity \quad (13)$$

$RootMessageProfile_{clu}$ represents the clustered form of the variables under the construct Root Message Profile, namely, Root Message Profile Clustered. *ProducerComInt* includes all the individual variables of the construct Producer Communicative Intention and their order-two interaction terms. *ProducerProfile* and *ProducerOpinionPolarity* include all the individual variables under the corresponding constructs, which are considered the control variables.

Table 22. Description of Clusters of Root Message Profile Clustered

Cluster Label	**Cluster Size**	**Cluster Characteristics**	**Avg. Text Quality Score**	**Avg. Sentiment-Emotion Score**	**Avg. Propagation Score**
0	298	All low	0.273	0.142	0.005
1	318	High Text Quality, High Sentiment-Emotion, Medium in Propagation	0.485	0.157	0.01
2	5	Medium in Text Quality, Medium in Sentiment-Emotion, High in Propagation	0.339	0.149	0.595

Table 22 describes the clusters computed for Root Message Profile Clustered over the misinformation root messages. The Text Quality Score, Sentiment-Emotion Score, and Propagation Score under Root Message Profile are considered as the three dimensions of each cluster. The clustering result by the K-Means algorithm (number of clusters set to 3) indicates that there are two main clusters (clusters 0 and 1) taking up almost all the root messages in the data set. According to the mean values of the three scores in each cluster, cluster 0 can be considered as representing those misinformation root messages (n = 298) that are relatively low in all the three scores. Cluster 1 represents those misinformation root messages (n = 318) that are high in their Text Quality Score and Sentiment-Emotion Score while medium in the Propagation Score.

Results of analyses in Table 23 reveal significant association between Producer Communicative Intention and Root Message Profile (supporting H3). It can be seen that, while all the variables under Production Intention and their interaction terms are significantly associated with the cluster assignment of Root Message Profile Clustered, they influence the

cluster assignment in different ways. With cluster 2 ignored due to its extremely small size (n = 5), the representative intention (prdInt_REP), directive intention (prdInt_DIR), expressive intention (prdInt_EXP), the count of intentions (prdInt_count), and most of the interaction term have strong, positive influence on the odds of being assigned to cluster 0. In comparison, the commissive intention (prdInt_COM) and quoting intention (prdInt_QOU) show positive influence on the odds for cluster 1.

Table 23. Regression of Producer Communicative Intention on Root Message Profile Clustered (H3)

IV's	Producer Communicative Intention	
DV	Root Message Profile Clustered	
Num. Obs.	590	
AIC	891.8287	
Cluster	#1	#2 (ignored)
Num. Sig. Main IV's	6	6
Num. Sig. Main IV's Inter.	15	15
Num. Sig. CV's	27	28
prdInt_REP	-1.71+	41.183+
prdInt_DIR	-38.345+	30.416+
prdInt_COM	37.149+	-1.997+
prdInt_EXP	-2.687+	-34.362+
prdInt_QOU	74.794+	-12.253+
prdInt_count	-28.578+	-44.093+
prdInt_REP * prdInt_DIR	-101.006+	-45.881+
prdInt_REP * prdInt_COM	-22.109+	-18.053+
prdInt_REP * prdInt_EXP	-68.277+	-16.178+
prdInt_REP * prdInt_QOU	9.07+	20.769+
prdInt_REP * prdInt_count	33.59+	2.261+
prdInt_DIR * prdInt_COM	-56.833+	15.841+
prdInt_DIR * prdInt_EXP	-100.38+	18.601+
prdInt_DIR * prdInt_QOU	-23.325+	-42.49+
prdInt_DIR * prdInt_count	67.481+	-2.551+
prdInt_COM * prdInt_EXP	-22.941+	8.362+
prdInt_COM * prdInt_QOU	54.814+	8.524+
prdInt_COM * prdInt_count	-9.919+	12.677+
prdInt_EXP * prdInt_QOU	9.994+	22.909+
prdInt_EXP * prdInt_count	33.331+	19.752+

prdInt_QOU * prdInt_count	-44.401+	-3.082+
opiPolClu1	-0.482+	57.081+
opiPolClu2	0.258+	-8.873+
p-value +: <0.05, ++: <0.01, +++: <0.001		

Then, the linear model in Equation (14) is established and fitted on the collection of misinformation root messages:

$$ConsumerProfile_{clu} = ProducerComInt + ProducerProfile + RootMessageProfile + ProducerOpinionPolarity \quad (14)$$

ConsumerProfile$_{clu}$ represents the clustered form of the variables under the construct Consumer Profile, namely, Consumer Clustered. *ProducerComInt* includes all the individual variables of the construct Producer Communicative Intention and their order-two interaction terms. *RootMessageProfile* and *ProducerOpinionPolarity* include all the individual variables under the corresponding constructs, which are considered the control variables.

Table 24. Description of Clusters of Consumer Profile Clustered

Cluster Label	**Cluster Size**	**Cluster Characteristics**	**Avg. Activeness Score**	**Avg. Socialization Score**	**Avg. Text Quality Score**	**Avg. Sentiment-Emotion Score**
0	102	High Text Quality, High Sentiment-Emotion	0.19	0.059	0.411	0.179
1	359	All med or low	0.224	0.056	0.286	0.101
2	155	High Activeness, High Socialization	0.365	0.109	0.387	0.143

Table 24 describes the clusters computed for Consumer Profile Clustered over the misinformation root messages. The Activeness Score, Socialization Score, Text Quality Score, and Sentiment-Emotion Score of the consumers associated with each root message are considered as the four dimensions of each cluster. The clustering result by the K-Means algorithm (number of clusters set to 3) indicates that there are three main clusters (clusters 0, 1, and 2) taking up all the root messages of misinformation. According to the mean values of the four scores in each cluster, cluster 0 can be considered as representing those misinformation root

messages (n = 102) that are mainly retweeted by consumers who are relatively high in their Text Quality Score and Sentiment-Emotion Score. Cluster 1 represents those misinformation root messages (n = 359) mainly retweeted by consumers who are relatively low or medium in all the scores. Cluster 2 represent those misinformation root messages (n = 155) mainly retweeted by consumers who are relatively high in their Activeness Score and Socialization Score.

Analytic results also show strong and significant influence of Producer Communicative Intention upon Consumer Profile (supporting H4). It can be seen in Table 25 that all the variables of Producer Communicative Intention and their interaction terms are significantly associated with the cluster assignment of Consumer Profile Clustered. For example, the representative, directive, and quoting intentions project strong, positive association the with odds of cluster 2. The commissive and expressive intentions present positive influence on the odds of cluster 0. Interaction terms of different intentions are positively associated with either cluster 1 or 2.

Table 25. Regression of Producer Communicative Intention on Consumer Profile Clustered (H4)

IV's	Producer Communicative Intention	
DV	Consumer Profile Clustered	
Num. Obs.	587	
AIC	1062.917	
Cluster	#1	#2
Num. Sig. Main IV's	6	6
Num. Sig. Main IV's Inter.	15	15
Num. Sig. CV's	51	51
prdInt_REP	-342.475+	422.702+
prdInt_DIR	-325.205+	13.932+
prdInt_COM	-324.524+	-442.207+
prdInt_EXP	-345.378+	-9.644+
prdInt_QOU	-342.669+	513.147+
prdInt_count	369.588+	25.658+
prdInt_REP * prdInt_DIR	65.098+	-34.401+
prdInt_REP * prdInt_COM	64.626+	423.021+
prdInt_REP * prdInt_EXP	45.416+	-58.735+
prdInt_REP * prdInt_QOU	43.216+	-580.385+
prdInt_REP * prdInt_count	-23.362+	74.262+
prdInt_DIR * prdInt_COM	85.627+	-138.521+
prdInt_DIR * prdInt_EXP	65.871+	58.888+
prdInt_DIR * prdInt_QOU	65.56+	58.956+
prdInt_DIR * prdInt_count	-43.05+	-41.147+
prdInt_COM * prdInt_EXP	65.789+	59.117+

prdInt_COM * prdInt_QOU	65.323+	57.408+
prdInt_COM * prdInt_count	-43.159+	-41.183+
prdInt_EXP * prdInt_QOU	44.83+	32.284+
prdInt_EXP * prdInt_count	-22.716+	-16.03+
prdInt_QOU * prdInt_count	-22.983+	-16.529+
opiPolClu1	0.12+	0.978+
opiPolClu2	0.312+	-0.71+
p-value +: <0.05, ++: <0.01, +++: <0.001		

3.5.3 Influence of Production Approach (H5 and H6)

First, the linear model in Equation (15) is established and fitted on the collection of misinformation root messages:

$$RootMessageProfile_{clu} = ProductionApproach + ProducerProfile + ProducerOpinionPolarity \quad (15)$$

$RootMessageProfile_{clu}$ represents the clustered form of the variables under the construct Root Message Profile, namely, Root Message Profile Clustered. *ProducerApproach* includes all the individual variables of the construct Producer Approach and their order-two interaction terms. *ProducerProfile* and *ProducerOpinionPolarity* include all the individual variables under the corresponding constructs, which are considered the control variables.

Results of analyses in Table 26 indicate strong influence of Production Approach upon Root Message Profile and Consumer Profile, suggesting that H5 and H6 are both supported. First, significant association is observed between all the variables under Production Approach (including their interaction terms) and the cluster assignment of Root Message Profile Clustered (in support of H5). Cluster 2 is ignored due to its small size (n = 5). The approaches of qualitative manipulation (pdtApp_QUA), relevance manipulation (pdtApp_REL), and several interaction terms are found to be positively associated with the odds of being assigned to cluster 1. The approaches of quantitative manipulation (pdtApp_QUT), image protection (pdtApp_IMG), and depersonalization (pdtApp_DEP), as well as the count of performed approaches are positively associated with the odds for cluster 0.

Table 26. Regression of Production Approach on Root Message Profile Clustered (H5)

IV's	Producer Approach	
DV	Root Message Profile Clustered	
Num. Obs.	590	
AIC	918.4405	
Cluster	#1	#2
Num. Sig. Main IV's	6	6
Num. Sig. Main IV's Inter.	15	15
Num. Sig. CV's	28	28
pdtApp_QUA	121.479+	65.59+
pdtApp_QUT	-31.064+	-56.086+
pdtApp_REL	96.828+	48.336+
pdtApp_IMG	-36.637+	-85.057+
pdtApp_DEP	-79.87+	20.053+
pdtApp_count	-36.935+	-11.565+
pdtApp_QUA * pdtApp_QUT	11.759+	14.045+
pdtApp_QUA * pdtApp_REL	136.625+	-48.73+
pdtApp_QUA * pdtApp_IMG	10.931+	8.436+
pdtApp_QUA * pdtApp_DEP	-32.217+	-39.506+
pdtApp_QUA * pdtApp_count	-83.124+	-11.817+
pdtApp_QUT * pdtApp_REL	-16.312+	-8.425+
pdtApp_QUT * pdtApp_IMG	-142.514+	68.96+
pdtApp_QUT * pdtApp_DEP	-184.995+	-24.799+
pdtApp_QUT * pdtApp_count	69.811+	4.419+
pdtApp_REL * pdtApp_IMG	-17.99+	11.024+
pdtApp_REL * pdtApp_DEP	-62.113+	-4.326+
pdtApp_REL * pdtApp_count	-55.316+	0.077+
pdtApp_IMG * pdtApp_DEP	-188.256+	37.692+
pdtApp_IMG * pdtApp_count	72.687+	25.657+
pdtApp_DEP * pdtApp_count	115.802+	-4.91+
opiPolClu1	-0.517+	28.738+
opiPolClu2	0.132+	-22.416+
p-value +: <0.05, ++: <0.01, +++: <0.001		

Furthermore, the linear model in Equation (16) is established and fitted on the collection of misinformation root messages:

$$ConsumerProfile_{clu} = ProductionApproach + ProducerProfile + RootMessageProfile + ProducerOpinionPolarity \quad (16)$$

ConsumerProfile$_{clu}$ represents the clustered form of the variables under the construct Consumer Profile, namely, Consumer Profile Clustered. *ProducerApproach* includes all the individual variables of the construct Producer Approach and their order-two interaction terms. *ProducerProfile*, *RootMessageProfile*, and ProducerOpinionPolarity include all the individual variables under the corresponding constructs, which are considered the control variables.

Results support H6 by showing that Production Approach is associated significantly with Consumer Profile. As shown in Table 27, all variables and their interaction terms under Production Approach are associated strongly with the cluster assignment of Consumer Profile Clustered. Specifically, use of the approaches of qualitative manipulation (pdtApp_QUA) and image protection (pdtApp_IMG) is associated positively with the odds for cluster 0. The approaches of quantitative manipulation (pdtApp_QUT), relevance manipulation (pdtApp_REL), and depersonalization (pdtApp_DEP) is found to be associated positively with the odds for cluster 1. The count of used approaches (pdtApp_count) is associated positively with the odds for cluster 2. Interaction terms of these individual variables are associated with different clusters labels.

Table 27. Regression of Production Approach on Consumer Profile Clustered (H6)

IV's	Production Approach	
DV	Consumer Profile Clustered	
Num. Obs.	587	
AIC	1069.402	
Cluster	#1	#2
Num. Sig. Main IV's	6	6
Num. Sig. Main IV's Inter.	15	15
Num. Sig. CV's	51	51
pdtApp_QUA	-461.455+	-113.951+
pdtApp_QUT	25.203+	-779.701+
pdtApp_REL	308.657+	-238.621+
pdtApp_IMG	-210.964+	-141.29+
pdtApp_DEP	291.724+	-480.705+
pdtApp_count	-24.018+	354.926+
pdtApp_QUA * pdtApp_QUT	-237.01+	65.517+
pdtApp_QUA * pdtApp_REL	-665.417+	-106.214+
pdtApp_QUA * pdtApp_IMG	-638.204+	539.469+
pdtApp_QUA * pdtApp_DEP	-140.249+	193.269+
pdtApp_QUA * pdtApp_count	403.324+	-323.989+

pdtApp_QUT * pdtApp_REL	95.79+	-496.69+
pdtApp_QUT * pdtApp_IMG	-151.176+	-126.074+
pdtApp_QUT * pdtApp_DEP	348.569+	-470.908+
pdtApp_QUT * pdtApp_count	-83.655+	341.746+
pdtApp_REL * pdtApp_IMG	-221.53+	62.532+
pdtApp_REL * pdtApp_DEP	278.394+	-284.45+
pdtApp_REL * pdtApp_count	-12.225+	155.365+
pdtApp_IMG * pdtApp_DEP	30.982+	87.222+
pdtApp_IMG * pdtApp_count	235.077+	-214.933+
pdtApp_DEP * pdtApp_count	-265.728+	128.082+
opiPolClu1	0.039+	0.881+
opiPolClu2	0.186+	-0.743+
p-value +: <0.05, ++: <0.01, +++: <0.001		

3.5.4 Influence of Producer Opinion Polarity (H7 and H8)

First, the linear model in Equation (17) is established and fitted on the collection of misinformation root messages:

$$RootMessageProfile_{clu} = ProducerOpinionPolarity + ProducerProfile \quad (17)$$

RootMessageProfile$_{clu}$ represents the clustered form of the variables under the construct Root Message Profile, namely, Root Message Profile Clustered. *ProducerOpinionPolarity* includes all the individual variables of the construct Production Opinion Polarity and their order-two interaction terms. *ProducerProfile* includes all the individual variables under the corresponding construct, which are considered the control variables.

Results of analyses indicate strong influence of Producer Opinion Polarity upon Root Message Profile and Consumer Profile, suggesting that H7 and H8 are both supported. First, significant association is observed in Table 28 between all the variables under Producer Opinion Polarity (including their interaction terms) and the cluster assignment of Root Message Profile Clustered (in support of H7).

Table 28. Regression of Producer Opinion Polarity on Root Message Profile Clustered (H7)

IV's	Producer Opinion Polarity
DV	Root Message Profile Clustered
Num. Obs.	590
AIC	899.9509
Cluster	#1

Num. Sig. Main IV's	2
Num. Sig. Main IV's Inter.	1
Num. Sig. CV's	26
rooPol	-1.517+
prdPol	-0.126+
rooPol * prdPol	-12.308+
p-value +: <0.05, ++: <0.01, +++: <0.001	

Furthermore, the linear model in Equation (18) is established and fitted on the collection of misinformation root messages:

$$ConsumerProfile_{clu} = ProducerOpinionPolarity + ProducerProfile + RootMessageProfile \quad (18)$$

ConsumerProfile$_{clu}$ represents the clustered form of the variables under the construct Consumer Profile, namely, Consumer Profile Clustered. *ProducerOpinionPolarity* includes all the individual variables of the construct Producer Opinion Polarity and their order-two interaction terms. *ProducerProfile* and *RootMessageProfile* include all the individual variables under the corresponding constructs, which are considered the control variables.

Results support H8 by showing that Producer Opinion Polarity is associated significantly with Consumer Profile. As shown in Table 29, all variables and their interaction terms under Producer Opinion Polarity are associated strongly with the cluster assignment of Consumer Profile Clustered.

Table 29. Regression of Producer Opinion Polarity on Consumer Profile Clustered (H8)

IV's	Producer Opinion Polarity	
DV	Consumer Profile Clustered	
Num. Obs.	587	
AIC	1038.306	
Cluster	#1	#2
Num. Sig. Main IV's	2	2
Num. Sig. Main IV's Inter.	1	1
Num. Sig. CV's	49	49

rooPol	-1.683+	7.078+
prdPol	-2.317+	-1.594+
rooPol * prdPol	-0.622+	-33.486+
p-value +: <0.05, ++: <0.01, +++: <0.001		

3.6 Discussion

In summary, all the hypotheses in Essay II have been supported based on results of analyses. In support of H1, observations indicate a strong association between Producer Opinion Polarity and Producer Communicative Intention. Results suggest that producers who are polarized over a subject (e.g., highly active towards COVID (high on rooPol)) in their root messages tend to be highly directive to the root message receivers (presenting the communicative intention of DIR) and moderately express their own emotion in the root messages the produce (presenting the communicative intention of EXP), namely, displaying highest odds for cluster 0. Producers who are highly emotional and sentimental in general tend to conceal their intention in the root messages, namely, trying to show as little communicative intention as possible (with highest odds for cluster 1). Producers who are simultaneously highly polarized in root messages (e.g., COVID-active) and highly sentimental and emotional in general also tend to be highly directive (DIR) and emotionally expressive (EXP) in their root messages.

The findings above suggest that producers' opinion polarity influences the communicative intention they hold when they produce misinformation. Both of producers' subject-independent polarity in general (as characterized by the sentiment and emotion expressed in their user description) and their polarity over a specific subject (as expressed in their root messages) contribute to this influence. In particular, the directive and expressive intention of the producers are driven by their opinion polarity most significantly.

Supporting H2, the results suggest that Producer Communicative Intention has significant influence on Production Approach. Specifically, if a producers hold strong representative intention (high in prdInt_REP) or quoting intention (high in prdInt_QUO) for their root messages, during the process of misinformation production, they tend to perform the approach of qualitative manipulation (namely, injecting completely fabricated story (QUA)) and, at the same time, try to disguise this by performing the approach of personal image protection (namely, to do anything to protecting their personal image (IMG)) and the approach of depersonalization

(namely, to distance themselves from the fake story (DEP)), presenting highest odds for cluster 2. Furthermore, if the producers hold strong directive intention (high in prdInt_DIR), commissive intention (high in prdInt_COM), or expressive intention (high in prdInt_EXP), they tend to perform the approach of quantitative manipulation (namely, providing authentic but only partial information (QUT)) (with highest odds for cluster 1).

The findings related to H2 suggest that producers' communicative intention influences the production approaches they apply when they produce misinformation. Furthermore, different types of communicative intentions can lead to the use of different types of production approaches. If the producers just intend to convey certain ideas to the receivers (e.g., with the communicative intention of REP or QUO), they choose to fabricate some falsehood (using QUA) and disguise their deceptive activities (using IMG and DEP). Alternatively, if the producers intend to motivate someone to take actions (e.g., with DIR or COM), they tend to use partially true information as an incentive (using QUT).

In support of H3, observations show that Producer Communicative Intention is significantly associated with Root Message Profile. Specifically, if the producers hold a commissive (prdInt_COM) or quoting intention (prdInt_QOU), the root messages they produce tend to present good text quality (with high Text Quality Score), be sentimental and emotional (with high Sentiment-Emotion Score), and be propagated broadly (with medium Propagation Score). In comparison, if they hold a representative (prdInt_REP), directive (prdInt_DIR), or expressive intention (prdInt_EXP), or two or more intentions simultaneously, their root messages tend to have low text quality (low Text Quality Score), low level of sentiment and emotion (low Sentiment-Emotion Score), and be propagated restrictively (low Propagation Score), namely, falling to cluster 1. These findings suggest that the communitive intentions held by producers can influence the root messages they produce. Depending on what types of communicative intentions they carry and express, the root message they produce present differences in text quality, sentiment and emotion, and strength of propagation.

As H4 is supported, the results indicate strong association between Producer Communicative Intention and Consumer Profile. For instance, producers with a representative (prdInt_REP), directive (prdInt_DIR), or quoting intention (prdInt_QOU) tend to be more attractive to consumers who are highly active (with high Activeness Score) and social (with high Socialization Score) on Twitter. Producers with a commissive (prdInt_COM) or expressive

intention (prdInt_EXP) tend to attract such consumers who present good text quality (with high Text Quality Score) and express more sentiment and emotion in their user description (with high Sentiment-Emotion Score). These finding indicate that the communicative intentions held by producers can impact which types of receivers are attracted to spread (retweet in this study) the root messages. Communicative intentions focusing on the receivers (e.g., REP, DIR, and QOU) tend to be attractive to more active and socialized receivers. In comparison, communicative intentions focusing on the producers themselves (e.g., COM and EXP) tend to be attractive to more sentimental or emotional receivers.

Supporting H5, it can be observed that Production Approach influences Root Message Profile significantly. Results suggest that if the producers insert purely fabricated story in the root messages (namely, performance the approach of qualitative manipulation (pdtApp_QUA)) or inflate the root messages with irrelevant details (namely, take the approach of relevance manipulation (pdtApp_REL)), the root messages they produce tend to have good text quality (with high Text Quality Score) and be highly sentimental and emotional, and can be propagated broader (with high Propagation Score). In comparison, if the producers make the root messages half-true half-false (performing the approach of quantitative manipulation (pdtApp_QUT)), or try make the root messages or themselves appear trustworthy (performing the approach of image protection (pdtApp_IMG)), or try to distance themselves from the root messages (performing the approach of depersonalization (pdtApp_DEP)), the root messages they produce then do have low text quality (with low Text Quality Score), low sentiment or emotion (with low Sentiment-Emotion Score), and low chance of being propagated (with low Propagation Score). The findings above suggest that which production approaches the producers choose can impact the root messages they produce. Using different production approaches can result in the production of distinctive root messages, which differ in their text quality, the sentiment and emotion expressed, and the propagation.

Furthermore, H6 is supported by the observation that Production Approach has strong, significant influence on Consumer Profile. The results suggest that if the producers (1) use purely fabricated stories in their root messages (namely, performing the approach of qualitative approach (pdtApp_QUA)), or (2) try to render the root messages authentic using linguistic methods (namely, performing the approach of image protection (pdtApp_IMG)), the root messages they produce tend to be most attractive to consumers who show good text quality

(meaning higher Text Quality Score) and more sentiment or emotion (meaning high Sentiment-Emotion Score) in their user description (cluster 0). When the approaches of quantitative manipulation (pdtApp_QUT), relevance manipulation (pdtApp_REL), or depersonalization (pdtApp_DEP) is performed during root message production, their root messages tend to attract consumer who are less active (with low Activeness Score) and socialized (low Socialization Score) on Twitter, and show lower text quality (lower Text Quality Score) and lower levels of sentiment and emotion (low Sentiment-Emotion Score) in their user description. These findings suggest that producers' choice of production approach can influence which types of receivers spread the root messages. Specifically, root messages produced with certain approaches tend to be more attractive to specific types of receivers. Root messages produced with fabricated information tend to be more attractive to receivers who are more sentimental or emotional. Root message produced with partially true information appears to be more attractive to less active and socialized receivers.

3.7 Conclusion

Essay II further examines the role of producers in the production and diffusion of misinformation on social media. The producers are characterized as social media users with their opinion polarity, their intention of producing misinformation, the approaches they take to producing misinformation. The relationships among these three features as well as their influences upon the root messages produced and consumers attracted are hypothesized and inspected. The analytic results drawn from a large, realistic data set have supported most of the hypothesized relationships and influences.

Opinion polarity of the producers regarding a specific subject (namely, opinion on the fight against the COVID-19 pandemic) is expressed as a function of (1) whether these producers are active or inactive towards the fight against COVID-19 and (2) the degree of their activeness and inactiveness. Their activeness or inactiveness are expressed in and measured from the root messages produced by these producers and the user description they composed.

How the hypotheses are supported by the analytic results is summarized in Figure 6. We can see that the producers' opinion polarity can influence the communicative intentions these producers hold when they produce misinformation—specifically, whether the producers intend to express their emotion and request the audience to do something (supporting H1). Further, their intentions can influence what approaches they take to producing the misinformation: do they

choose to inject purely fabricated stories in the root messages or partially false stories? Do they try to convince the readers that these stories are authentic? Do they try to conceal that they are the authors? (supporting H2)

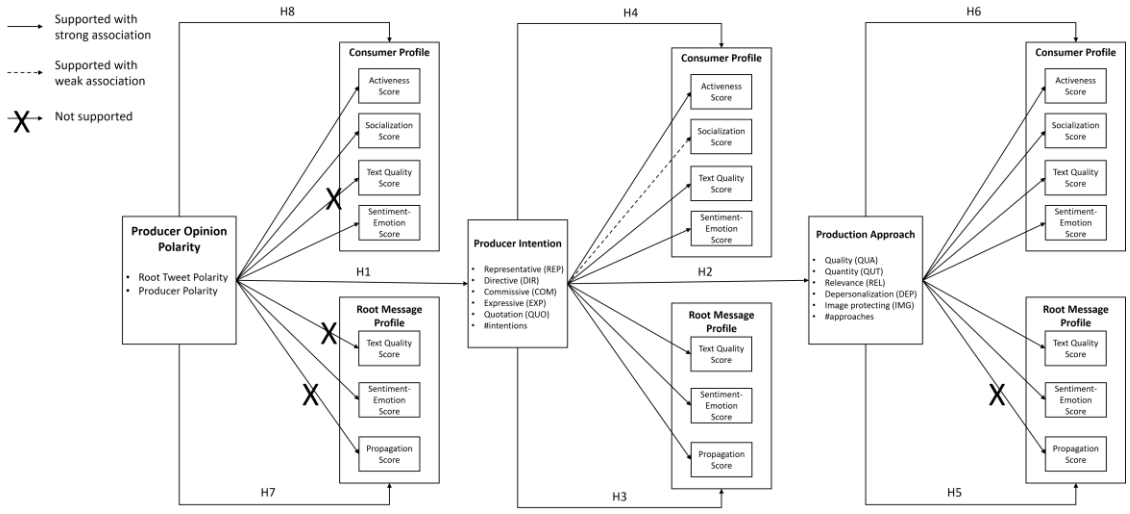

Figure 6. Support to Hypotheses in Essay II

Furthermore, intentions of the producers (especially the intentions for conveying some "facts", requesting the receivers or the producers themselves to do something, expressing certain emotion, and transmitting someone else's words) held by the producers can influence the root messages they produce in terms of the text quality, degree of sentiment and emotion, and propagation of the root messages (supporting H3). With different intentions, producers can attract different types of users to retweet, who are differentiated by their degree of activeness and socialization on Twitter, and the text quality, sentiment, and emotion they express in their user description (supporting H4).

At the same time, the approaches taken by the producers to producing the root messages (e.g., do they completely or partially fabricate the stories? Do they do anything in the root messages to conceal their authorship and the falsehood of the stories?) can influence the text quality, sentiment and emotion, and propagation of the root messages they produce (supporting H5). Producers' choice of production approaches can also influence the which types of users are attracted to retweet (supporting H6).

Finally, opinion polarity of the producers (e.g., are they strongly inclined to a specific opinion on a given subject, say, the fight against COVID-19? Are they highly emotional and

sentiment in general, independent from any subject?) can influence the text quality, sentiment and emotion, and propagation of the root messages they produce (supporting H7). Producers' opinion polarity can also influence the which types of users are most attracted to retweet (supporting H6).

To sum up, on the basis of the previous essay, Essay II further fills the second research gap that the producers and their roles in misinformation production a diffusion have not been clarified. Furthermore, Essay II fills the third and fourth research gaps by recognizing three new features of producers—their opinion polarity, their communicative intention, and the production approach they choose—and examines how these features influence the root messages and consumers. All the findings together produce a picture of the misinformation producers as humans engaged in language production.

CHAPTER IV: ESSAY III: A MACHINE LEARNING FRAMEWORK FOR MISINFORMATION DETECTION

4.1 Introduction

Essay III builds upon Essay I and II and aims at filling the fifth research gap. In this part, a framework for mitigating the damage of misinformation is proposed, implemented, and evaluated. This framework leverages predictive modeling based on machine learning in order to detect which messages contain misinformation.

In combat against misinformation, main platforms such as Facebook and Twitter have started publicly marking the posts suspicious of containing misinformation as false information (BBC News, 2020; Sophie Lewis, 2020). In research literature, various features have been used to train machine learning classifiers in order to detect misinformation (e.g., Karimi et al., 2018; Reis et al., 2019a; Shu, Sliva, et al., 2017; Shu, Wang, et al., 2019; X. Zhou et al., 2019). In one representative approach (Reis et al., 2019a), a wide range of features related to the root messages are used, such as syntactic (e.g., n-gram and readability) and lexical characteristics (e.g., number of words and unique words, first-person pronouns, and hashtags) of the root messages. Some features of the sources (i.e., producers) are also considered, such as credibility and trustworthiness of the sources, and the number of friends, followers, and likes of the producers. In another representative approach (Shu, Wang, et al., 2019), the relation between producers and new stories and the relation between consumers and new stories are used as features. Specifically, which users have produced root messages with a specific new story and which users have spread this news story are considered features.

Compared to the existing solutions, the proposed framework utilizes a wider range of knowledge to improve the performance of misinformation detection. The input includes knowledge of producers obtained in Essay I and II—e.g., knowledge extracted from their online profiles, their communicative intentions, and their polarity of opinion—as well as knowledge about consumers and root messages. The framework's performance has been examined experimentally under various configurations with respect to input features and algorithms.

4.2 Theoretical Background and Literature Review

4.2.1 Predictive Modeling in Misinformation Research

Essay III follows the structures of prior works in Information Systems on predictive modeling (e.g., Chau et al. 2020; He et al. 2018; Liebman et al. 2019; Zhang and Ram 2020) and proposes a framework for detecting root messages with misinformation out of non-misinformation root messages. The proposed framework utilizes the technology of predictive modeling. Predictive modeling is a process that uses data and computational methods to construct mathematical models and uses these models to predict outcomes of events. These models typically take the data about real world as input and estimate the consequences or outcomes as output. Techniques such as statistical modeling and machine learning are used to construct the predictive models (Kuhn et al., 2013; Parish & Duraisamy, 2016; Richiardi et al., 2013). Significant research has been conducted to leverage predictive modeling for combating misinformation on social media.

Seeing the harm incurred by misinformation, especially the social disturbance during and after the 2016 election, researchers in various fields have been motivated intensely to create techniques based on predictive modeling in order to detect misinformation or fake news on social media. The existing solutions consist of two broad categories, including the news content-based methods and social content-based methods (Shu, Sliva, et al., 2017). The news content-based methods are based on human intuition. These methods estimate the veracity of root messages simply based on the content of the messages. Various linguistic clues are extracted from the message bodies and titles (Aldwairi & Alwahedi, 2018; Pérez-Rosas et al., 2017), such as n-grams, punctuation, psycholinguistic features (e.g., emotion, sentiment, use of verbs and nouns), and message readability (e.g., number of characters, complex words, readability indices (Kincaid et al., 1975)). Supervised machine learning serves as the dominating type of algorithms to differentiate between misinformation and true information (Shu, Sliva, et al., 2017). Typically, these algorithms require that root messages, which is unstructured data, be converted into vectors representing the key terms used in the text, which is structured data (Ozbay & Alatas, 2020). Then, vectorized representations of the root messages together with their veracity pre-determined as ground truth are sent to supervised machine learning algorithms, such as neural network and support vector machine. These algorithms build predictive models and then estimate the veracity for root messages the ground truth of which is unknown.

However, pure content-based methods are generally not effective because misinformation, such as fake news stories, is often intentionally composed to mislead users by mimicking true information. They bear similar linguistic characteristics as true messages (Shu, Wang, et al., 2019). To solve this problem, researchers proposed social content-based methods for misinformation detection.

Social-content based methods typically leverage the knowledge about the propagation patterns of root messages, such as their propagation paths (Yang Liu & Wu, 2018; Wu & Liu, 2018). Various kinds of additional information, such as user profiles, user activities, and user social networks (Monti et al., 2019), can be incorporated into the representation of message propagation patterns, forming networks of knowledge. Such networks can also be enriched by the knowledge about the relationship among producers, root messages, and consumers (Shu, Wang, et al., 2019). Altogether, these knowledge (or a apart of them) can enable the construction of predictive models that estimate the veracity of root messages (Yang Liu & Wu, 2018; Monti et al., 2019; Wu & Liu, 2018). In addition to supervised machine learning, unsupervised algorithms, such as collapsed Gibbs sampling (S. Yang et al., 2019), have been found efficient in supporting the model construction process.

In addition, efforts also have been done to predict the potential consumers (also called victims or susceptible users) of misinformation, namely, receivers who spread the root messages further. Knowledge about receivers not only enables us to prevent social media users from being victimized by misinformation directly, it has also been integrated into larger information systems for detecting misinformation (Boshmaf, Logothetis, et al., 2015). The risk for receivers to become consumers of misinformation depends on a combination of their account-related, linguistic, network-related, and behavioral features, which can be used to build accurate predictive models using machine learning techniques (Guerra et al., 2013; Shen et al., 2019; C. Wagner et al., 2012). Account features refer to account metadata such as number of followers, friends, posts, and comments, account age, and user description of the receivers (Shen et al., 2019). Linguistic features refer to the receivers' use of specific keywords in the misinformation messages indicating, for example, positive or negative emotions, socialization activities (C. Wagner et al., 2012), and perception and cognitive thinking (Shen et al., 2019). Network features characterize these receivers' followers and friends in the same social network, including measures about how much they are clustered in the network (Shen et al., 2019), and how closely

they are connected to highly socialized users (Guerra et al., 2013; C. Wagner et al., 2012). Behavioral features describe how often the receivers make conversation, if they can make use a diversity of words and topics, etc. (Guerra et al., 2013; C. Wagner et al., 2012) Characteristics of misinformation receivers have also been explored from the psychological perspective (Pennycook & Rand, 2020). It is found that people who tend to ascribe profundity to randomly generated sentences, who overclaim their level of knowledge, and who are strong in analytical thinking are more likely to consume misinformation such as fake news and bullshit.

4.2.2 Detection of Misinformation using Predictive Modeling

A significant research gap can be identified from the current literature in the detection of misinformation messages. So far, the existing solutions mainly rely on information about the receivers as well as a limited set of linguistic features of the root messages, such as the appearance of particular keywords. While these kinds of information are important for characterizing root messages' risk of containing misinformation, the knowledge about misinformation producers has not been fully utilized. As one of the three key components of misinformation diffusion, producers decide how misinformation is produced; their decisions on misinformation production can also influence how misinformation will be consumed.

4.3 Design of Framework

To address the research gap discussed above, I propose the design of a new framework for detecting root messages with misinformation. The central task of the framework is to detect misinformation on social media, namely, to estimate *if a given root message contains misinformation*. As illustrated in Figure 7, the framework executes a workflow with four phases to support this task, including Model Construction, Data Preparation, Risk Estimation, and Warning.

Compared to the existing solutions to this problem (e.g., Boshmaf, Logothetis, et al. 2015; Boshmaf, Ripeanu, et al. 2015; Guerra et al. 2013; Shen et al. 2019; Wagner et al. 2012), the proposed framework not only uses a broad part of online profiles of producers and consumers as input, but also makes use of another two types of knowledge mined in Essay I and II from the data, including (1) producers' communicative intentions, and their opinion polarity, and (2) characteristics of root messages, their content, and their propagation.

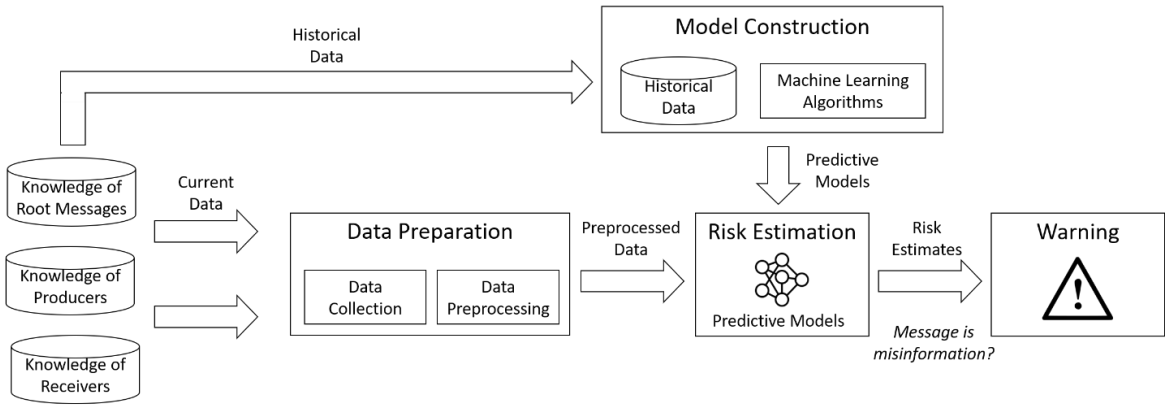

Figure 7. Workflow of Misinformation Detection Framework

Model Construction

This phase is executed asynchronously in relation to the other three phases. In this phase, a historical collection of data is used, which consists of (1) knowledge of a large number of historical misinformation root messages (e.g., the content of these messages, their linguistic features, emotions, sentiments, and topics), (2) knowledge of producers of the root messages (e.g., their account profiles, their social connections, their opinion polarity regarding related issues, and their use of speech acts in the root messages, and (3) knowledge of consumers of the root messages (i.e., all the features in their profile).

Selected machine learning algorithms are executed on the historical data in order to train machine learning models that can predict root messages' risk of containing misinformation. These predictive models are sent to the Risk Estimation phase. The database of historical data and the model training processes are updated when new data is collected during Data Preparation or the performance of Risk Estimation varies.

Data Preparation

First, knowledge of all the producers, root message, and consumers is collected from the platform. Then, all these data are preprocessed to make them ready to serve the predictive models. Preprocessed data is copied to the database in Model Construction.

Risk Estimation

In this phase, preprocessed data is fed on the predictive models in order to estimate if root messages contain misinformation. Performance of the prediction is collected later and fed back to Model Construction, in order to improve the model training processes.

Warning

Finally, warnings are sent to the receivers of the root messages. The receivers' reactions to the warnings and the root message (e.g., reply to the root message, retweet the root message, or ignore the warning) are collected in order to compute the effectiveness of the entire system.

The framework can also help to discover important features of producers, consumers, and root messages to effectively detect misinformation root messages. Prior work (W. Zhang & Ram, 2020) has shown that performance of a predictive framework can be used as evaluator of feature importance.

4.4 Methodology

Essay III proposes a framework for detecting misinformation on social media. It presents an improved, predictive modeling-based solution to the problem of detecting root messages containing misinformation. Adding to the existing solutions discussed, the proposed framework makes use of (1) the online profiles of producers, root messages, and consumers of misinformation, and (2) novel producer-related features, such as polarity of their opinions and their communicative intentions.

I designed and prototyped the proposed framework and inspect its performance experimentally, in order to verify if including the aforementioned knowledge about producers, root messages, and consumers can improve the performance of risk estimation. A variety of configurations of the system were tested, such as different prediction algorithms and feature selection methods. Algorithmically, multiple, existing supervised machine learning methods, such as neural network, support vector machine, and decision tree (Qiu et al., 2016) were benchmarked in order to test the robustness of the design with different algorithms. Furthermore, prediction performance of all the configurations above were compared and analyzed based on two model evaluation metrics, including True Positive Rate and Area Under ROC Curve (AUC) (Lobo et al., 2008; Shaheen & Myers, 2007).

The prototype was implemented on a HP laptop with Inter i7 Duo Core 2.7 of GHz and 16 GB memory. The data preprocessing procedures were mainly implemented using the Dataframe operations and text processing functions in NLTK (v3.5) on Python. In terms of predictive modeling, the supervised classifiers were implemented using the Scikit-learn library (v0.0) on Python. The visualization of evaluation results was produced using the Seaborn (v0.11) library on Python. All the activities above were performed on Python version 3.8.2.

4.5 Evaluation and Results

4.5.1 Process of Evaluation

A prototype of the misinformation detection system was built and evaluated. The prototype uses machine learning classifiers to predict if a given root message contains misinformation ("1") as labeled by the fact checking websites or not ("0"). In total, seven different types of classifiers were tested, including artificial neural network (ANN), decision tree (DT), Gaussian Naive Bayes (GNB), k-nearest neighbors (KNN), logistic regression (LG), random forests (RF), and support vector machine (SVM). The hyperparameters of these algorithms were configured using the library-default values. Values of the selected key hyperparameters are listed in Table 30.

Table 30. Values of Selected Hyperparameters

ANN	DT	GNB	KNN	LR	RF	SVM
n_layers: 1 hidden_layer_sizes: 10 learning_rate_init: 0.001 max_iter: 500 momentum: 0.9	Criterion: gini min_samples_split: 2 min_samples_leaf: 1	No hyperpar. supported	n_neighbors: 5	fit_intercept: TRUE max_iter: 100	n_estimators: 100 criterion: gini min_samples_leaf: 1	kernel: rbf strength_regularization: 1

The algorithms were tested with different input data sets, each of which was a combination of the following five base data sets, including: the variables under Producer Profile (denoted as Pro, see Table 1), Root Message Profile (denoted as Roo, see Table 3), Consumer Profile (denoted as Ret, see Table 2), Producer Communicative Intention (denoted as Int, see Table 16), and Producer Opinion Polarity (denoted as Pol, including Root Message Polarity and Producer Polarity). For instance, if an input data set include the variables from Producer Profile and Producer Communicative Intention, this data set is denoted as Pro+Int. Individual features included these feature sets are summarized in Table 31.

Table 31. Input Feature Sets of Prototype

Feature Set Name	Feature Category	Feature Name
Producer Profile (Pro)	Account Metainfo	Followers count, Friends count, Listed count, Likes count, Posts count, Geo-enabled, Verified
	Sentiment and Emotion	Emotion_fear, Emotion_anger, Emotion_sad, Emotion_disgust, Emotion_surprise, Emotion_anticipation, Emotion_trust, Emotion_joy, Emotion_positive, Emotion_negative, Sentiment_positive, Sentiment_neutral, Sentiment_negative, Sentiment_compound
	Text Quality	Subjectivity, Readability, Characters count, Words count, Unique words count, Sentences count, Characters per word, Words per sentence

Root Message Profile (Roo)	Text Quality	Subjectivity, Readability, Characters count, Words count, Unique words count, Sentences count, Characters per word, Words per sentence
	Sentiment and Emotion	Emotion_fear, Emotion_anger, Emotion_sad, Emotion_disgust, Emotion_surprise, Emotion_anticipation, Emotion_trust, Emotion_joy, Emotion_positive, Emotion_negative, Sentiment_positive, Sentiment_neutral, Sentiment_negative, Sentiment_compound
	Propagation	Cascade size, Likes count
Consumer Profile (Ret) *Mean and media of each variable is included*	Activeness	Account age, Likes count, Posts count
	Socialization	Followers count, Friends count, Listed count, Percentage geo-enabled consumers, Percentage verified consumers, Percentage protected consumers
	Text Quality	Subjectivity, Readability, Characters count, Words count, Unique words count, Sentences count, Characters per word, Words per sentence
	Sentiment and Emotion	Emotion_fear, Emotion_anger, Emotion_sad, Emotion_disgust, Emotion_surprise, Emotion_anticipation, Emotion_trust, Emotion_joy, Emotion_positive, Emotion_negative, Sentiment_positive, Sentiment_neutral, Sentiment_negative, Sentiment_compound
Producer Communicative Intention (Int)		REP, DIR, COM, EXP, QOU, Count of communicative intentions
Producer Opinion Polarity (Pol)		Root Message Polarity, Producer Polarity

For each unique pair of algorithm and input feature set, a ten-fold cross validation was performed ten times on the entire data set which is randomly shuffled. That means, each pair of algorithm and feature data set was tested 10X10 times in the evaluation.

Finally, while multiple performance metrics were used in the evaluation, this essay focuses on the True Positive Rate (TPR) and Area under the ROC (AUC). TPR is especially important for information detection systems as it measures the proportion of misinformation messages a system can detects, reflecting the system's capability in detecting misinformation as much as possible. However, a system with high TPR might also suffer from a high False Positive Rate (FPR) by raising positive alarms excessively. AUC incorporates both TPR and FPR and evaluates the performance of a system in a more balanced manner.

4.5.2 Performance of Misinformation Detection System

All the evaluation results of the prototype are present in Table 32 and Table 33. First, the performance of the three basic feature sets—Pro, Roo, and Ret—was evaluated. The mean TPR and AUC over the 100 runs are presented in Figure 8 and Figure 9. Overall, we can see that the system is able to produce a TPR above 0.8 and AUC above 0.65 with most of the feature sets and algorithms. The maximum TPR of 0.94 is achieved by SVM. A high AUC close to 0.85 is achieved by both SVM and RF.

Table 32. Performance of Prediction – True Positive Rate (TPR)

Input Feature Set	Num. Features	ANN	DT	GNB	KNN	LR	RF	SVM
Pro	29	0.816	0.744	0.728	0.834	0.847	0.894	0.899
Roo	24	0.838	0.701	0.810	0.824	0.895	0.907	0.945
Ret	59	0.812	0.735	0.486	0.783	0.840	0.877	0.871
Pro+Roo+Ret	112	0.800	0.752	0.534	0.831	0.822	0.887	0.887
Int	6	0.990	0.965	0.523	0.821	1.000	0.969	0.971
Pol	2	0.817	0.752	0.806	0.832	0.821	0.812	0.816
Int+Pol	8	0.857	0.759	0.821	0.859	0.851	0.841	0.872
Pro+Int	35	0.797	0.743	0.719	0.845	0.851	0.902	0.908
Pro+Pol	31	0.851	0.783	0.755	0.864	0.863	0.898	0.906
Pro+Int+Pol	37	0.827	0.781	0.769	0.871	0.864	0.907	0.901
Roo+Int	30	0.813	0.706	0.715	0.797	0.875	0.899	0.936
Roo+Pol	26	0.841	0.758	0.791	0.834	0.848	0.870	0.875
Roo+Int+Pol	32	0.844	0.759	0.797	0.827	0.859	0.887	0.893
Ret+Int	65	0.809	0.739	0.513	0.773	0.833	0.874	0.865
Ret+Pol	61	0.834	0.773	0.516	0.802	0.835	0.885	0.854
Ret+Int+Pol	67	0.828	0.774	0.534	0.800	0.837	0.881	0.852
Pro+Roo+Ret+Int	118	0.796	0.743	0.549	0.820	0.820	0.886	0.890
Pro+Roo+Ret+Pol	114	0.809	0.792	0.568	0.848	0.837	0.892	0.879
Pro+Roo+Ret+Int+Pol	120	0.814	0.793	0.585	0.836	0.830	0.890	0.885

Table 33. Performance of Prediction – Area Under ROC Curve (AUC)

Input Feature Set	Num. Features	ANN	DT	GNB	KNN	LR	RF	SVM
Pro	29	0.688	0.631	0.716	0.664	0.713	0.761	0.714
Roo	24	0.657	0.555	0.640	0.657	0.657	0.678	0.687
Ret	59	0.764	0.610	0.716	0.733	0.762	0.787	0.785
Pro+Roo+Ret	112	0.761	0.634	0.733	0.733	0.765	0.816	0.796
Int	6	0.650	0.650	0.637	0.581	0.625	0.651	0.538
Pol	2	0.784	0.642	0.783	0.736	0.778	0.753	0.758
Int+Pol	8	0.803	0.653	0.772	0.761	0.800	0.766	0.799
Pro+Int	35	0.689	0.628	0.727	0.666	0.723	0.766	0.719
Pro+Pol	31	0.789	0.687	0.774	0.735	0.810	0.832	0.816
Pro+Int+Pol	37	0.777	0.688	0.777	0.728	0.814	0.831	0.814
Roo+Int	30	0.679	0.557	0.683	0.659	0.696	0.698	0.713
Roo+Pol	26	0.776	0.655	0.772	0.745	0.788	0.819	0.801
Roo+Int+Pol	32	0.788	0.653	0.774	0.752	0.800	0.827	0.817
Ret+Int	65	0.766	0.612	0.725	0.744	0.763	0.788	0.790
Ret+Pol	61	0.797	0.655	0.740	0.781	0.808	0.822	0.821

Ret+Int+Pol	67	0.798	0.662	0.747	0.785	0.810	0.825	0.827
Pro+Roo+Ret+Int	118	0.763	0.625	0.739	0.742	0.768	0.817	0.799
Pro+Roo+Ret+Pol	114	0.781	0.686	0.751	0.768	0.805	0.843	0.826
Pro+Roo+Ret+Int+Pol	120	0.790	0.689	0.757	0.767	0.805	0.844	0.831

More importantly, the four feature sets produce differential performance. For almost all the algorithms, Roo alone leads to the maximum TPR. Adding more features by using Pro+Roo+Ret does not improve does not maximize the TPR. However, Roo produces the lowest AUC, while Pro+Roo+Ret produces the maximum AUC for most of the algorithms.

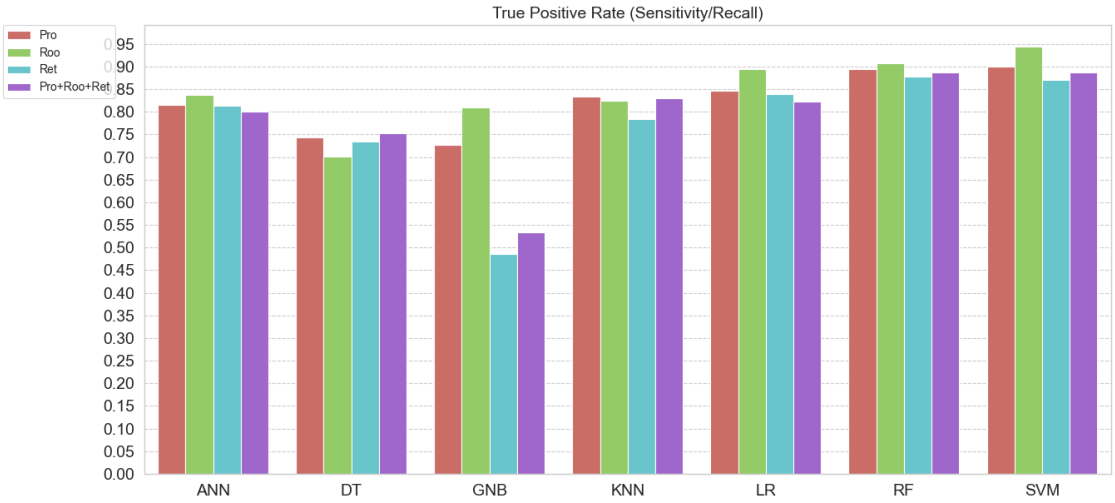

Figure 8. Performance of Producer Profile, Root Message Profile, and Consumer Profile (TPR)

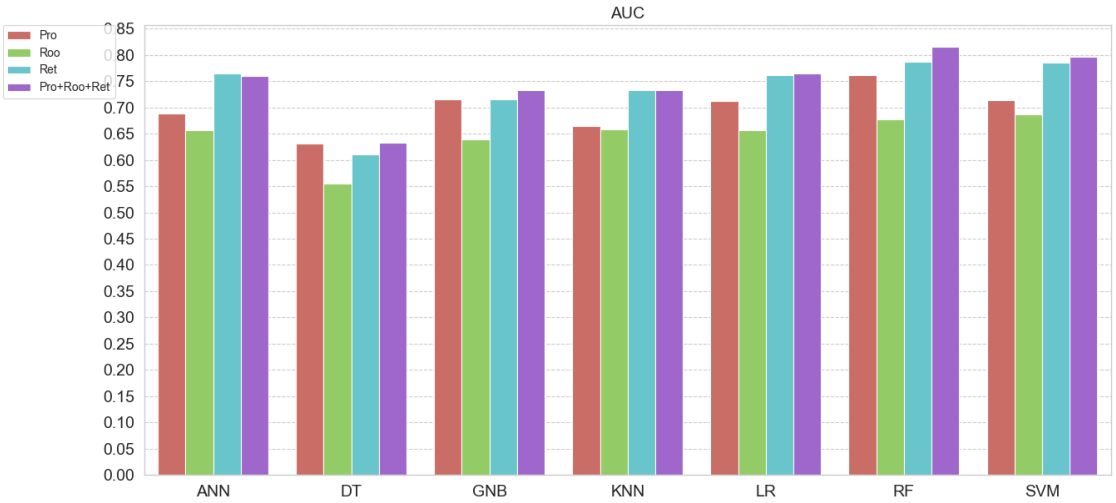

Figure 9. Performance of Producer Profile, Root Message Profile, and Consumer Profile (AUC)

4.5.3 Influence of Producer Communicative Intention and Producer Opinion Polarity

In the next step, the impact of Producer Communicative Intention and Producer Opinion Polarity upon performance of prediction was evaluated systematically. To this end, Int and Pol were added on top of Pro, Roo, and Ret for evaluation. First, we can see in Figure 10 and Figure 11 that both TPR and AUC of Producer Profile are improved when Producer Communicative Intention and Producer Opinion Polarity were added. For multiple algorithms (e.g., DT, KNN, LR, and RF with TPR and DT, GNB, LR, and RF with AUC), Pol introduces a stronger improvement than Int, while adding both of them maximizes the performance.

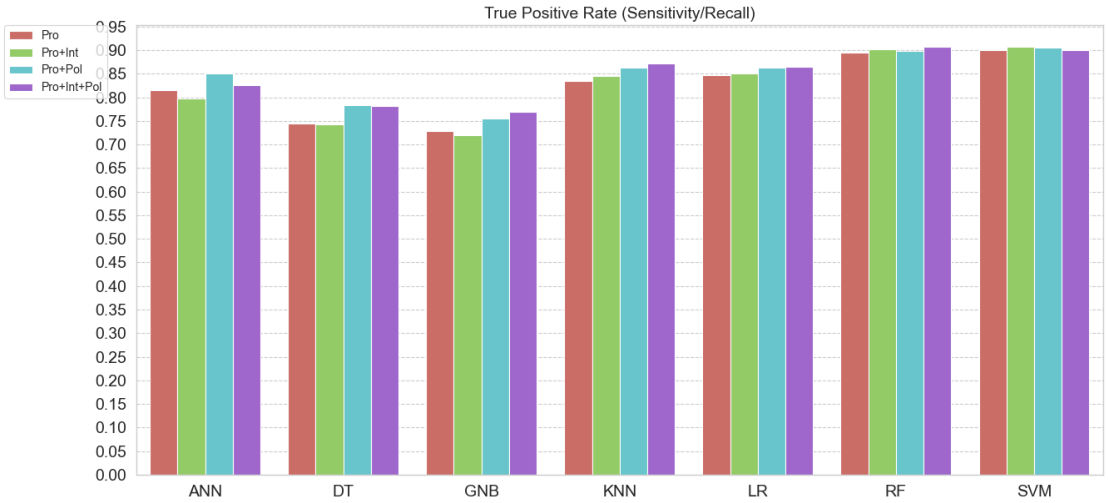

Figure 10. Performance of Producer Profile with Producer Communicative Intention and Producer Opinion Polarity (TPR)

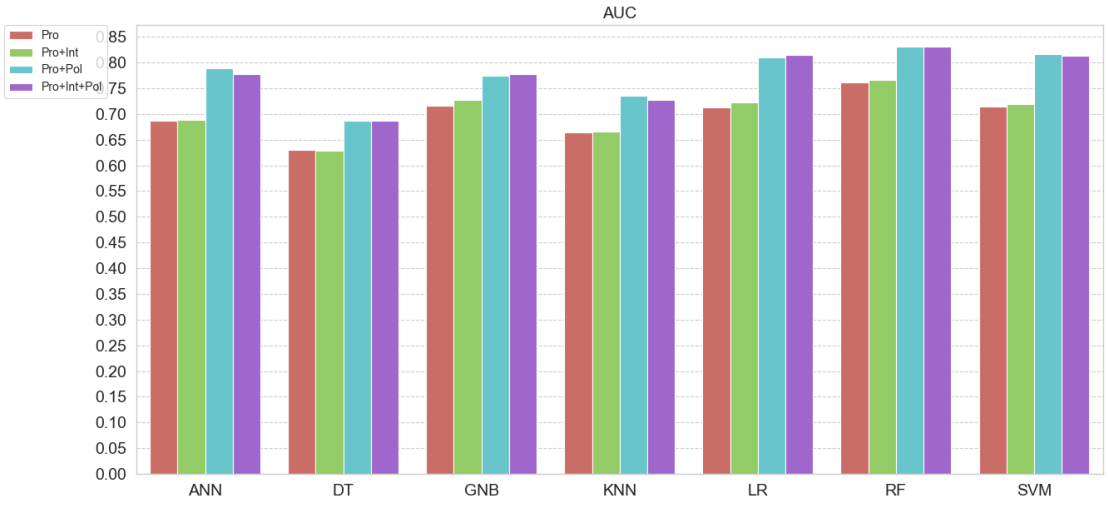

Figure 11. Performance of Producer Profile with Producer Communicative Intention and Producer Opinion Polarity (AUC)

A similar pattern of improvement from Int and Pol is generally available for Consumer Profile (see Figure 14 and Figure 15) and Pro+Roo+Ret (see Figure 16 and Figure 17). The most significant exception, however, is with Root Message Profile in TPR (see Figure 12). We can see that adding Pol and, especially, Int lowers the TPR for ANN, KNN, LR, RF, and SVM.

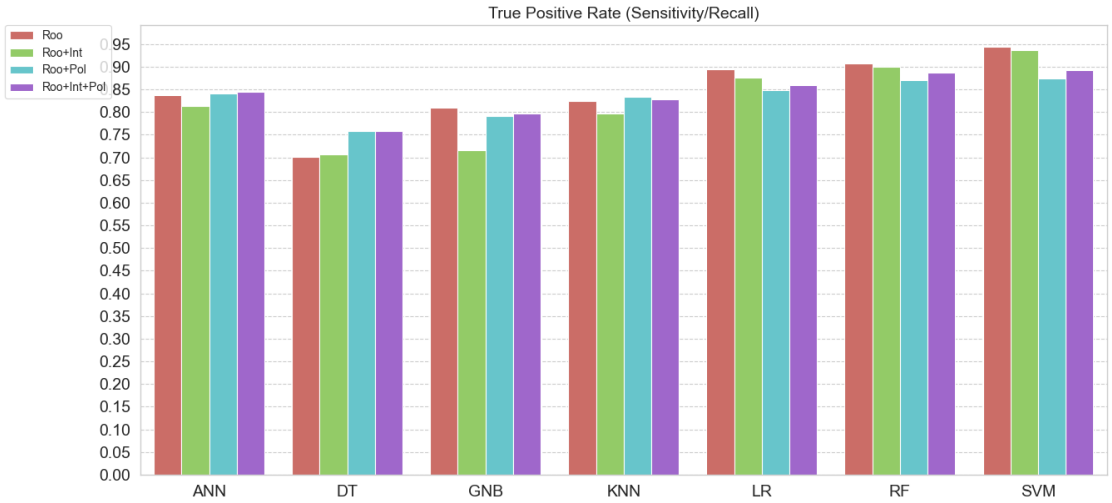

Figure 12. Performance of Root Message Profile with Producer Communicative Intention and Producer Opinion Polarity (TPR)

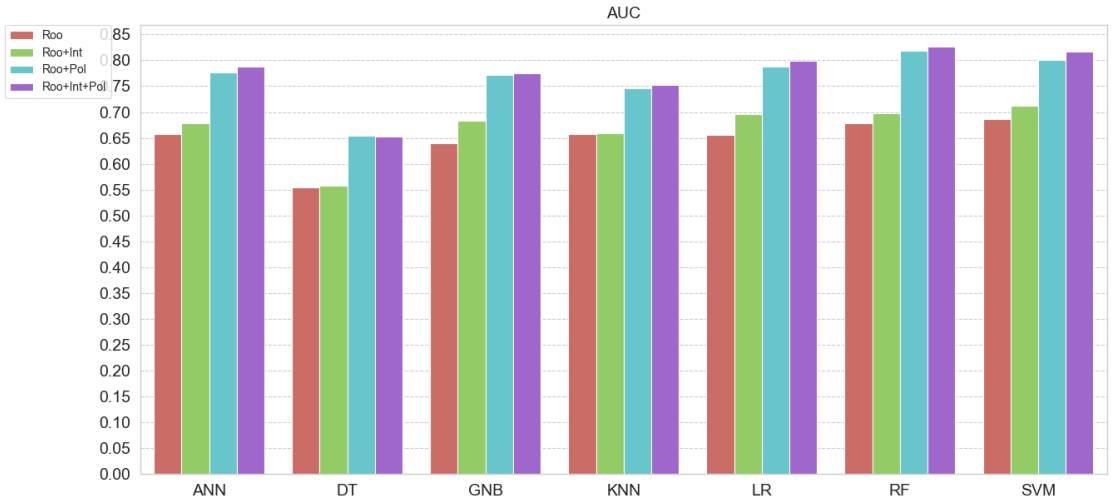

Figure 13. Performance of Root Message Profile with Producer Communicative Intention and Producer Opinion Polarity (AUC)

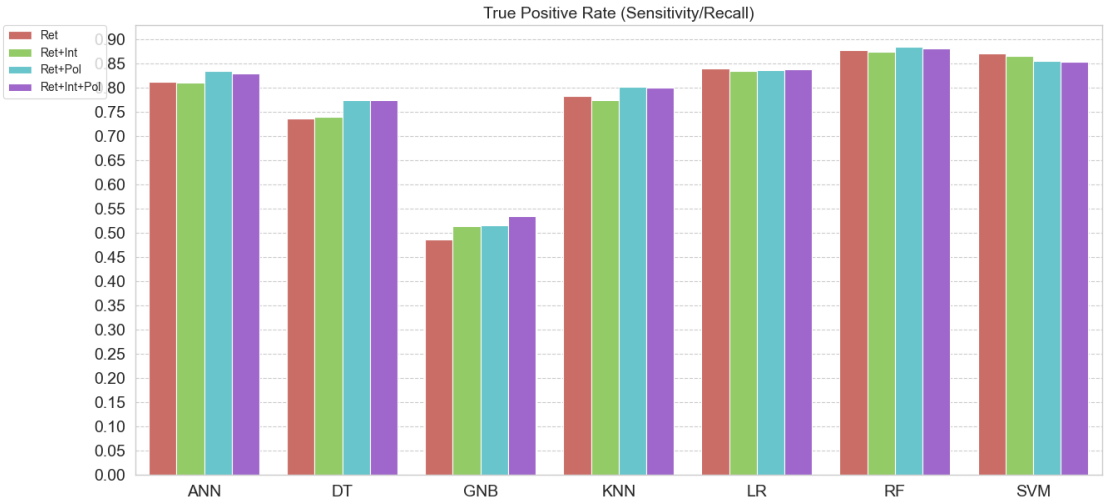

Figure 14. Performance of Consumer Profile with Producer Communicative Intention and Producer Opinion Polarity (TPR)

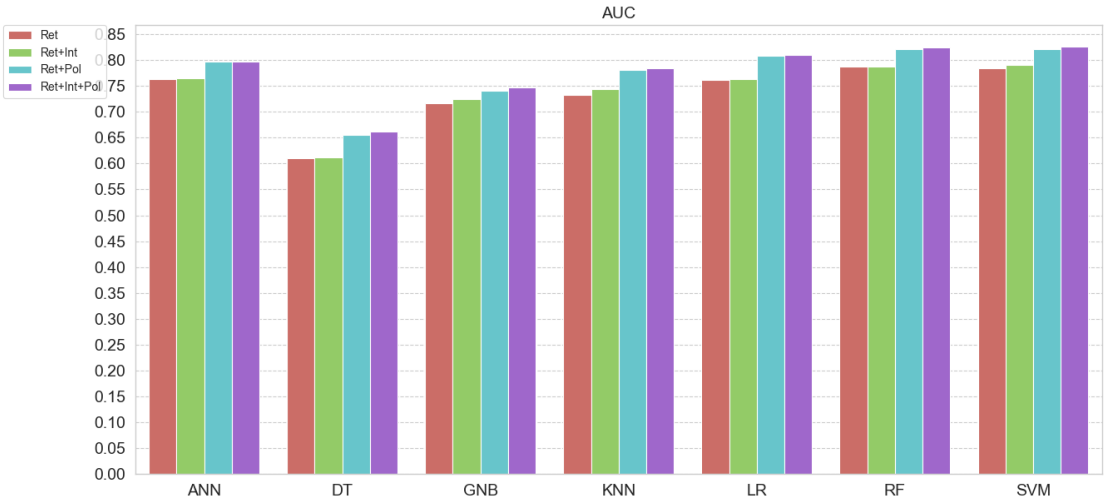

Figure 15. Performance of Consumer Profile with Producer Communicative Intention and Producer Opinion Polarity (AUC)

Figure 16. Performance of All Profiles with Producer Communicative Intention and Producer Opinion Polarity (TPR)

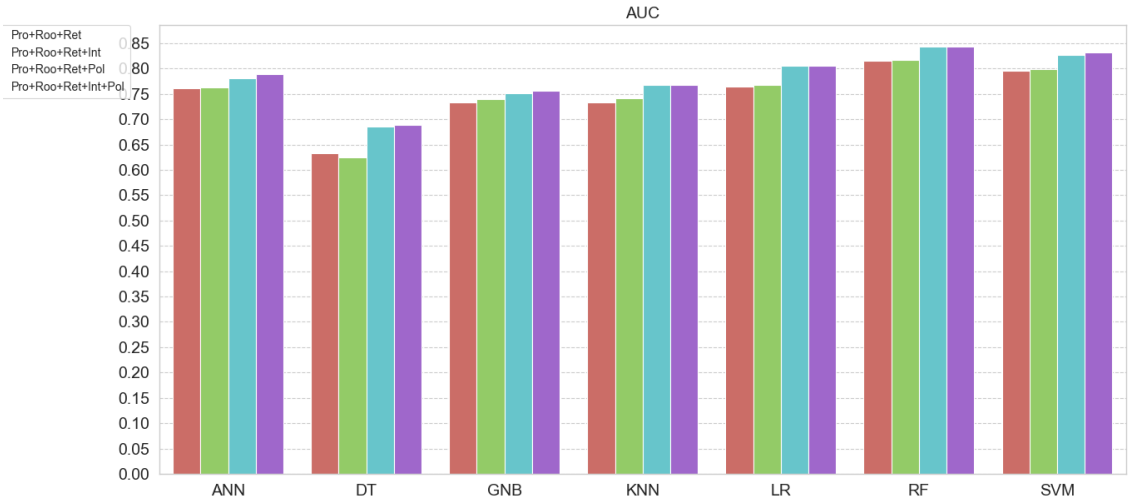

Figure 17. Performance of All Profiles with Producer Communicative Intention and Producer Opinion Polarity (AUC)

4.6 Discussion

The results of the evaluation reveal several phenomena. First and foremost, the system prototype can deliver a strong prediction performance by learning the knowledge about the producers, root messages, and consumers. Online profiles of these three components of misinformation alone can be used to drive machine learning algorithm to achieve strong capability in detecting misinformation.

Secondly, knowledge about producers' intention and their opinion polarity can help improve the system even further on top of the online profiles. However, the online profiles and knowledge about intention/opinion polarity need to be set up in particular ways. For instance, some combination of intention/opinion polarity and online profiles might even undermine the performance.

Furthermore, multiple factors moderate the effects of the knowledge about producers' intention and opinion polarity. For instance, intention shows differential effects across different algorithms; when the underlying profiles are changed, the effects of intention might be reverted.

The findings presented above suggest that the knowledge extracted from the online profiles of misinformation producers, root messages, and consumers can enhance the detection of misinformation using the machine learning technology. Given that these types of knowledge can be easily obtained by social media platforms, this study provides motivation for the platforms to incorporate the knowledge into their automatic mechanism of misinformation mitigation.

Furthermore, since the major social platforms have access to unlimited computing resources and huge volumes of producers, root messages, and consumers, they are able to fine-tune the online profiles in real time and always trace the subset of profiles with the greatest predictive power for their prediction tasks. That means the methods proposed in this essay might become more effective if deployed to these large platforms.

4.7 Conclusion

In this essay, I designed and prototyped a machine learning framework to detect root messages containing misinformation. This framework is novel in that it utilizes a broad range of knowledge to support the machine learning process, including the online profiles of producers, root messages, consumers, and the communication intention and opinion polarity of the producers. While such an extensive range of user and content profiles has not been used in the

detection of misinformation, the features from Producer Communicative Intention and Producer Opinion Polarity are the first time to be extracted and applied in any research.

Evaluation results demonstrate that, overall, the system has achieved substantially high performance in misinformation detection. Furthermore, smartly combining the features from Producer Profile, Root Message Profile, and Consumer Profile can improve the performance. Finally, introducing features from Producer Opinion Polarity and Producer Communicative Intention can further enhance the performance.

In summary, Essay III fills the fifth research gap that the exiting solutions to detecting misinformation rely on knowledge of the root messages and consumers while underutilizing knowledge of the producers; in addition, only a limited subset of the online profiles have been utilized. The proposed system is able to incorporate knowledge of all the three components of misinformation. Features contained in the online profiles directly or features needed to be extracted by comprehending the pragmatics and semantics of the online profiles are both prepared and provided to the system. Evaluation results indicate that the two types of features above need to be combined together in order to achieve maximized performance of misinformation detection.

CHAPTER V: IMPLICATIONS, CONTRIBUTIONS, LIMITATIONS AND FURTURE RESEARCH

Essays I and II provide implications for both research and practice. On the research side, it fills several gaps in the literature of social media study. First, compared to the consumers and root messages, producers of misinformation have attracted rather limited research attention. In this study, a systematic examination of producers of misinformation is conducted. Their role in the process of misinformation diffusion is clarified. In particular, interactions between producers and two important components of misinformation diffusion are explored, which have not been extensively studied before. These interactions include: the one between producers and message propagation, and between producers and consumers.

Second, little efforts have been made so far to scrutinize the producers as human communicators conducting language production, asking what they want behind the text when they are engaged in misinformation production. My study will fill this gap by exploring their communicative intentions and speech acts. This would make the current research one of the pioneer studies exploring the pragmatic forces that drive misinformation production and how these forces are passed on to misinformation propagation and consumption.

In practice, this research contributes deepens our understanding of misinformation diffusion. With the knowledge about producers' communicative intentions and production approaches, people now have a clearer picture of what these producers want and what they would do realize it. Based on these understandings, new information systems can be built to detect misinformation producers. Knowledge about producers' behavior patterns will be helpful for maximizing the performance of such systems.

Further, the role of producers' opinion polarization in the entire process of misinformation diffusion is examined. Doing this could renew our understanding to several important social media phenomena, such as echo chamber, bubble filtering, and confirmation bias. Their causes, principles, and effects will be reexamined now from the perspective of opinion polarization of producers.

In Essay III, an extended subset of the online profiles of the producers, root messages, and consumers is used in detection of misinformation. This extended subset also includes the knowledge of producer opinions and their communicative intentions. In practice, experience

from Essay III can be used to build a new generation of information systems for mitigating misinformation. One the one hand, since an extended set of features (i.e., predictors) are available, the new generation of misinformation mitigating system is able to tune the feature set used to support the current prediction task continuously and in real time. The tuning process consists of incorporating features from the total collection of online profiles and excluding features from the feature set in use, with the goal of finding out the feature set with maximized predictive power.

On the other hand, thanks to the availability of the extended subset of online profiles, the new generation of misinformation mitigating system is able to support multiple prediction tasks (i.e., multiple prediction goals), including detecting the misinformation root messages, detecting the potential producers of misinformation, and detecting the potential consumers of misinformation. Each of these tasks can be employed by the system to mitigate the damage of misinformation. In turn, an overall performance goal is set up by the social platform to measure the ultimate efficiency of the system to mitigate the damage of misinformation. The platform or the system itself can determine which task to switch to under and overall performance goal and the current workloads.

The main contributions of the dissertation lie in three dimensions. In the first dimension, the dissertation fills four significant gaps in the research of misinformation on social media. First, the dissertation examines and clarifies the roles of producers, root messages, and consumers in the production and diffusion of misinformation based on their online profiles. Second, the role of the producers in misinformation production and diffusion is examined and clarified. Their effects on the root messages and consumers are inspected. Third, producers' opinion polarity, communicative intention, and the production approach they use are examined individually and integratively; these three features represent the inspection of producers on the ideological, pragmatic, and semantic-syntactic level, respectively. Fourth, the influences of these three features of producers on the root messages and consumers are scrutinized.

In the second dimension, the dissertation makes contributions in extending the usability of two classic linguistic theories, the Speech Act Theory (SAT) and Interpersonal Deception Theory (IDT). Traditionally, the SAT has been used to inspect speech acts in text. In this dissertation, SAT is used innovatively to inspect communicative intention of misinformation producers. Furthermore, in literature, IDT has been mostly used to understand off-line

communication. In the current study, IDT serves for the first time to examine how producers produce misinformation on social media. On the one hand, introducing these theories into the current study strengthens the theoretical foundation of the dissertation. On the other hand, to my knowledge, both theories are applied in the area of misinformation for the first time. It is a contribution for the current study to assign new applicable areas to both classic theories.

In the third dimension, the dissertation makes contributions by filling an additional gap in practice by proposing a new machine learning-based software framework for detecting misinformation on social media. Extending the existing solutions, the proposed framework uses a much broader scope of knowledge of the producers, root messages, and consumers; the effects (both positive and negative) of each of these three sources of knowledge on the performance of misinformation detection is analyzed. Using the extended collection of predictors for detecting misinformation not only improves the performance of misinformation detection, but also creates the possibility of building a new generation of misinformation mitigation systems: such a system can freely tune the set of predictors and switch among multiple prediction targets, such as predicting the misinformation root messages, producers, and consumers.

The main limitation of the dissertation is that the annotation of communicative intention and production approach were only conducted by the candidate alone. The annotation of more than 900 root messages took about 72 hours or 9 workdays, not including the 1-2 hours spent in getting familiar with the annotation methods. In particular, the annotation of production approaches was especially time consuming, as the annotator had to read the full articles on the fact-checking websites to determine if each root message used quantitative (QUA) or qualitative (QUT) manipulation. Therefore, I was not able to find anyone willing to take such a large workload. In addition, due to the difficulty in having face-to-face meetings during the pandemic, I was not able to familiarize the potential annotators with the annotating methods efficiently.

Another limitation of the dissertation is that the sample size of root message is relatively low. The data collection is a highly time-consuming task, which needs to be repeated everyday ideally. I was not able to start collecting data in January 2020 when COVID-19 started to spread. In addition, due to the fact that I was not able to involve other people into the project, I only have the capacity to collect root messages from four fact-checking websites.

In addition, the current study is limited in that all the hypotheses were built only upon the misinformation root messages. I have not examined if the findings are valid in non-

misinformation or authentic root messages. I also did not examine if there exist any new relationships among the producers, root messages, and consumers in the authentic root messages.

In the future, the work in the dissertation can be improved or extended in several areas. First, additional annotators can be involved to improve the quality of data annotation. Employing multiple annotators, while more difficult than using a single annotator, provides a higher degree of methodological rigor, and thus lending a higher degree of confidence in the findings (Boyer & Verma, 2000). The crowdsourcing services such as Mechanic Turk or Survey Monkey can be employed in the future to bring in annotations from multiple people.

Secondly, the data collection needs to be extended until the end of the COVID-19 pandemic. Increasing the sample size can not only improve the confidence in the current findings, but also creates opportunity of discovering new patterns and trends in the data. Furthermore, other types of data related to the root messages can also be included and examined, such as replies, posts made by the producers and consumers unrelated to the root messages, and profiles of non-consumers of the root messages.

Third, regression models in Essay I and II can be further simplified in order to make the results more interpretable. Currently, some models include many single terms and interaction terms, which increase the risk of collinearity (Midi et al., 2010). In the future, robust tests (Olea & Pflueger, 2013) can be performed to alleviate this problem. Alternatively, stepwise model reduction can be performed to exclude the insignificant terms (Peduzzi et al., 1980).

Fourth, in the future, the misinformation and authentic root messages can be both analyzed and compared. It would be interesting to inspect if the research models validated for misinformation data is still valid for authentic data. New relationships among the producers, root messages, and consumers might be discovered by comparing the results produced on root messages with different veracity.

Sixth, the current software framework focuses on the detection of misinformation. By extending the training and evaluation data, some new prediction tasks can be supported by the framework, such as predicting the risk of receivers to become consumers, and estimating the risk of any producers to be misinformation producers.

REFERENCES

Abbas, A., Zhou, Y., Deng, S., & Zhang, P. (2018). Text analytics to support sense-making in social media: A language-action perspective. *MIS Quarterly*, *42*(2).

Abd-Alrazaq, A., Alhuwail, D., Househ, M., Hamdi, M., & Shah, Z. (2020). Top concerns of tweeters during the COVID-19 pandemic: infoveillance study. *Journal of Medical Internet Research*, *22*(4), e19016.

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

Ajao, O., Bhowmik, D., & Zargari, S. (2019). Sentiment aware fake news detection on online social networks. *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2507–2511.

Al-Rawi, A. (2019). Gatekeeping fake news discourses on mainstream media versus social media. *Social Science Computer Review*, *37*(6), 687–704.

Alamsyah, A., & Adityawarman, F. (2017). Hybrid sentiment and network analysis of social opinion polarization. *2017 5th International Conference on Information and Communication Technology (ICoIC7)*, 1–6.

Albright, L., Cohen, A. I., Malloy, T. E., Christ, T., & Bromgard, G. (2004). Judgments of communicative intent in conversation. *Journal of Experimental Social Psychology*, *40*(3), 290–302.

Aldwairi, M., & Alwahedi, A. (2018). Detecting fake news in social media networks. *Procedia Computer Science*, *141*, 215–222.

Alhidari, A., Iyer, P., & Paswan, A. (2015). Personal level antecedents of eWOM and purchase intention, on social networking sites. *Journal of Customer Behaviour*, *14*(2), 107–125.

Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, *31*(2), 211–236.

Anstead, N., & O'Loughlin, B. (2015). Social media analysis and public opinion: The 2010 UK general election. *Journal of Computer-Mediated Communication*, *20*(2), 204–220.

Arielli, E. (2018). Sharing as speech act. *Versus*, *47*(2), 243–258.

Au, C. H., Ho, K. K. W., & Chiu, D. K. W. (2021). The Role of Online Misinformation and Fake News in Ideological Polarization: Barriers, Catalysts, and Implications. *Information Systems Frontiers*, 1–24.

Austin, J. L. (1975). *How to do things with words* (Vol. 88). Oxford university press.

Awan, I. (2017). Cyber-extremism: Isis and the power of social media. *Society*, *54*(2), 138–149.

Bach, K. (n.d.). *SPEECH ACTS*. Routledge Encyclopedia of Philosophy. http://userwww.sfsu.edu/kbach/spchacts.html

Badaro, G., Jundi, H., Hajj, H., El-Hajj, W., & Habash, N. (2018). Arsel: A large scale arabic sentiment and emotion lexicon. *OSACT*, *3*, 26.

Bagozzi, R. P., Dholakia, U. M., & Pearo, L. R. K. (2007). Antecedents and consequences of online social interactions. *Media Psychology*, *9*(1), 77–114.

Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. B. F., Lee, J., Mann, M., Merhout, F., & Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*, *115*(37), 9216–9221.

Baldwin, T., Cook, P., Lui, M., MacKinlay, A., & Wang, L. (2013). How noisy social media text, how diffrnt social media sources? *Proceedings of the Sixth International Joint Conference on Natural Language Processing*, 356–364.

Bandari, R., Asur, S., & Huberman, B. (2012). The Pulse of News in Social Media: Forecasting Popularity. *ICWSM 2012 - Proceedings of the 6th International AAAI Conference on Weblogs and Social Media*.

Banja, J. D. (2010). Intentions, deception, and motivated reasoning. *AJOB Neuroscience*, *1*(1), 65–66.

Barberá, P. (2014). How social media reduces mass political polarization. Evidence from Germany, Spain, and the US. *Job Market Paper, New York University*, *46*.

Bavelas, J. B., Black, A., Chovil, N., & Mullett, J. (1990). *Equivocal communication.* Sage Publications, Inc.

BBC News. (2020). *Coronavirus: Twitter will label Covid-19 fake news.* BBC News. https://www.bbc.com/news/technology-52632909

Beam, M. A., Hutchens, M. J., & Hmielowski, J. D. (2018). Facebook news and (de) polarization: reinforcing spirals in the 2016 US election. *Information, Communication &*

Society, *21*(7), 940–958.

Bessi, A., Petroni, F., Del Vicario, M., Zollo, F., Anagnostopoulos, A., Scala, A., Caldarelli, G., & Quattrociocchi, W. (2015). Viral misinformation: The role of homophily and polarization. *Proceedings of the 24th International Conference on World Wide Web*, 355–356.

Bessi, A., Zollo, F., Del Vicario, M., Puliga, M., Scala, A., Caldarelli, G., Uzzi, B., & Quattrociocchi, W. (2016). Users polarization on Facebook and Youtube. *PloS One*, *11*(8), e0159641.

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

Bondielli, A., & Marcelloni, F. (2019). A survey on fake news and rumour detection techniques. *Information Sciences*, *497*, 38–55.

Borella, C. A., & Rossinelli, D. (2017). *Fake News, Immigration, and Opinion Polarization*.

Boshmaf, Y., Logothetis, D., Siganos, G., Ler\'\ia, J., Lorenzo, J., Ripeanu, M., & Beznosov, K. (2015). Integro: Leveraging Victim Prediction for Robust Fake Account Detection in OSNs. *Ndss*, *15*, 8–11.

Boshmaf, Y., Ripeanu, M., Beznosov, K., & Santos-Neto, E. (2015). Thwarting fake OSN accounts by predicting their victims. *Proceedings of the 8th ACM Workshop on Artificial Intelligence and Security*, 81–89.

Bovet, A., & Makse, H. A. (2019). Influence of fake news in Twitter during the 2016 US presidential election. *Nature Communications*, *10*(1), 1–14.

Bowers, J. W., Elliott, N., & Desmond, R. J. (1977). Exploiting pragmatic rules: Devious messages. *Human Communication Research*, *3*(3), 235–242.

Boxell, L. (2020). Demographic change and political polarization in the United States. *Economics Letters*, 109187.

Boxell, L., Gentzkow, M., & Shapiro, J. M. (2017). *Is the internet causing political polarization? Evidence from demographics*.

Boyer, K. K., & Verma, R. (2000). Multiple raters in survey-based operations management research: a review and tutorial. *Production and Operations Management*, *9*(2), 128–140.

Bradshaw, S., & Howard, P. N. (2018). Challenging truth and trust: A global inventory of organized social media manipulation. *The Computational Propaganda Project*, *1*.

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

Brennen, J. S., Simon, F., Howard, P. N., & Nielsen, R. K. (2020). Types, sources, and claims of Covid-19 misinformation. *Reuters Institute*, *7*.

Buller, D. B., & Burgoon, J. K. (1996). Interpersonal deception theory. *Communication Theory*, *6*(3), 203–242.

Buller, D. B., Burgoon, J. K., Daly, J. A., & Wiemann, J. M. (1994). Deception: Strategic and nonstrategic communication. *Strategic Interpersonal Communication*, 191–223.

Burgoon, J. K., & Buller, D. B. (1994). Interpersonal deception: III. Effects of deceit on perceived communication and nonverbal behavior dynamics. *Journal of Nonverbal Behavior*, *18*(2), 155–156.

Burgoon, J. K., Buller, D. B., Guerrero, L. K., Afifi, W. A., & Feldman, C. M. (1996). Interpersonal deception: XII. Information management dimensions underlying deceptive and truthful messages. *Communications Monographs*, *63*(1), 50–69.

Calvillo, D. P., Ross, B. J., Garcia, R. J. B., Smelter, T. J., & Rutchick, A. M. (2020). Political ideology predicts perceptions of the threat of covid-19 (and susceptibility to fake news about it). *Social Psychological and Personality Science*, *11*(8), 1119–1128.

Carr, C. T., Schrock, D. B., & Dauterman, P. (2012). Speech acts within Facebook status messages. *Journal of Language and Social Psychology*, *31*(2), 176–196.

Chamberlain, P. R. (2010). Twitter as a Vector for Disinformation. *Journal of Information Warfare*, *9*(1), 11–17.

Chandler, J. D., Salvador, R., & Kim, Y. (2018). Language, brand and speech acts on Twitter. *Journal of Product & Brand Management*.

Chau, M., Li, T. M. H., Wong, P. W. C., Xu, J. J., Yip, P. S. F., & Chen, H. (2020). Finding People with Emotional Distress in Online Social Media: A Design Combining Machine Learning and Rule-Based Classification. *MIS Quarterly*, *44*(2).

Chen, T., Li, Q., Yang, J., Cong, G., & Li, G. (2019). Modeling of the public opinion polarization process with the considerations of individual heterogeneity and dynamic conformity. *Mathematics*, *7*(10), 917.

Chen, Z., Liu, B., Hsu, M., Castellanos, M., & Ghosh, R. (2013). Identifying intention posts in discussion forums. *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 1041–1050.

Chiou, L., & Tucker, C. (2018). *Fake news and advertising on social media: A study of the anti-vaccination movement*.

Chou, W.-Y. S., Oh, A., & Klein, W. M. P. (2018). Addressing health-related misinformation on social media. *Jama*, *320*(23), 2417–2418.

Choudhary, A., & Arora, A. (2021). Linguistic feature based learning model for fake news detection and classification. *Expert Systems with Applications*, *169*, 114171.

Cinelli, M., Quattrociocchi, W., Galeazzi, A., Valensise, C. M., Brugnoli, E., Schmidt, A. L., Zola, P., Zollo, F., & Scala, A. (2020). The covid-19 social media infodemic. *ArXiv Preprint ArXiv:2003.05004*.

Coggins, T. E., & Carpenter, R. L. (1981). The communicative intention inventory: A system for observing and coding children's early intentional communication. *Applied Psycholinguistics*, *2*(3), 235–251.

Colliander, J. (2019). "This is fake news": Investigating the role of conformity to other users' views when commenting on and spreading disinformation in social media. *Computers in Human Behavior*, *97*, 202–215.

Conover, M. D., Ratkiewicz, J., Francisco, M., Gonçalves, B., Menczer, F., & Flammini, A. (2011). Political polarization on twitter. *Fifth International AAAI Conference on Weblogs and Social Media*.

Cuerie, H. C. (1952). A projection of socio-linguistics: The relationship of speech to social status. *Southern Journal of Communication*, *18*(1), 28–37.

Danescu-Niculescu-Mizil, C., Gamon, M., & Dumais, S. (2011). Mark my words! Linguistic style accommodation in social media. *Proceedings of the 20th International Conference on World Wide Web*, 745–754.

De Choudhury, M., Counts, S., & Horvitz, E. (2013). Predicting postpartum changes in emotion and behavior via social media. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 3267–3276.

de Oliveira, M. J., Huertas, M. K. Z., & Lin, Z. (2016). Factors driving young users' engagement with Facebook: Evidence from Brazil. *Computers in Human Behavior*, *54*, 54–61.

de Waal, A., & Ibreck, R. (2013). Hybrid social movements in Africa. *Journal of Contemporary African Studies*, *31*(2), 303–324. https://doi.org/10.1080/02589001.2013.781320

Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., Stanley, H. E., & Quattrociocchi, W. (2016). The spreading of misinformation online. *Proceedings of the National Academy of Sciences*, *113*(3), 554–559.

Del Vicario, M., Vivaldo, G., Bessi, A., Zollo, F., Scala, A., Caldarelli, G., & Quattrociocchi, W. (2016). Echo chambers: Emotional contagion and group polarization on facebook. *Scientific Reports*, *6*, 37825.

DiMaggio, P., Evans, J., & Bryson, B. (1996). Have American's social attitudes become more polarized? *American Journal of Sociology*, *102*(3), 690–755.

Ding, X., Liu, T., Duan, J., & Nie, J.-Y. (2015). Mining User Consumption Intention from Social Media Using Domain Adaptive Convolutional Neural Network. *AAAI*, *15*, 2389–2395.

Ebert, F., Castor, F., Novielli, N., & Serebrenik, A. (2018). Communicative intention in code review questions. *2018 IEEE International Conference on Software Maintenance and Evolution (ICSME)*, 519–523.

Edmond, C. (2013). Information manipulation, coordination, and regime change. *Review of Economic Studies*, *80*(4), 1422–1458.

Er\csahin, B., Akta\cs, Ö., K\il\inç, D., & Akyol, C. (2017). Twitter fake account detection. *2017 International Conference on Computer Science and Engineering (UBMK)*, 388–392.

Evans, D. A. (2016). *Situations and Speech Acts: toward a formal semantics of discourse*. Routledge.

Falk, E. B., Morelli, S. A., Welborn, B. L., Dambacher, K., & Lieberman, M. D. (2013). Creating buzz: the neural correlates of effective message propagation. *Psychological Science*, *24*(7), 1234–1242.

Falk, E., O'Donnell, M. B., & Lieberman, M. D. (2012). Getting the word out: neural correlates of enthusiastic message propagation. *Frontiers in Human Neuroscience*, *6*, 313.

Ferrara, E. (2015a). " Manipulation and abuse on social media" by Emilio Ferrara with Ching-man Au Yeung as coordinator. *ACM SIGWEB Newsletter*, *Spring*, 1–9.

Ferrara, E. (2015b). Manipulation and abuse on social media. *ACM SIGWEB Newsletter*, *Spring*, 1–9.

Ferrara, E. (2017). Disinformation and social bot operations in the run up to the 2017 French presidential election. *ArXiv Preprint ArXiv:1707.00086*.

Fishbein, M., & Ajzen, I. (1977). *Belief, attitude, intention, and behavior: An introduction to theory and research*.

Fitch, K. L., & Sanders, R. E. (2004). *Handbook of language and social interaction*. Psychology Press.

Flintham, M., Karner, C., Bachour, K., Creswick, H., Gupta, N., & Moran, S. (2018). Falling for fake news: investigating the consumption of news via social media. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–10.

Frangonikolopoulos, C. A., & Chapsos, I. (2012). Explaining the role and the impact of the social media in the Arab Spring. *Global Media Journal: Mediterranean Edition*, *7*(2).

Garcia, D., Mendez, F., Serdült, U., & Schweitzer, F. (2012). Political polarization and popularity in online participatory media: an integrated approach. *Proceedings of the First Edition Workshop on Politics, Elections and Data*, 3–10.

Garimella, V. R. K., & Weber, I. (2017). A long-term analysis of polarization on Twitter. *Eleventh International AAAI Conference on Web and Social Media*.

Georgiou, N., Delfabbro, P., & Balzan, R. (2020). COVID-19-related conspiracy beliefs and their relationship with perceived stress and pre-existing conspiracy beliefs. *Personality and Individual Differences*, 110201.

Gift, K., & Gift, T. (2015). Does politics influence hiring? Evidence from a randomized experiment. *Political Behavior*, *37*(3), 653–675.

Gikas, J., & Grant, M. M. (2013). Mobile computing devices in higher education: Student perspectives on learning with cellphones, smartphones & social media. *The Internet and Higher Education*, *19*, 18–26.

Goel, S., Watts, D. J., & Goldstein, D. G. (2012). The structure of online diffusion networks. *Proceedings of the 13th ACM Conference on Electronic Commerce*, 623–638.

Goldman, E., & Slezak, S. L. (2006). An equilibrium model of incentive contracts in the presence of information manipulation. *Journal of Financial Economics*, *80*(3), 603–626.

Gollmann, D. (2012). Veracity, plausibility, and reputation. *IFIP International Workshop on Information Security Theory and Practice*, 20–28.

Granik, M., & Mesyura, V. (2017). Fake news detection using naive Bayes classifier. *2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON)*, 900–903.

Grice, H. P. (1957). Meaning. *The Philosophical Review*, 377–388.

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

Grundlingh, L. (2018). Memes as speech acts. *Social Semiotics*, *28*(2), 147–168.

Gruzd, A., & Roy, J. (2014). Investigating political polarization on Twitter: A Canadian perspective. *Policy & Internet*, *6*(1), 28–45.

Guerra, P. C., Meira Jr, W., Cardie, C., & Kleinberg, R. (2013). A measure of polarization on social media networks based on community boundaries. *Seventh International AAAI Conference on Weblogs and Social Media*.

Guy Rosen. (2020). *An Update on Our Work to Keep People Informed and Limit Misinformation About COVID-19*. Facebook. https://about.fb.com/news/2020/04/covid-19-misinfo-update/

H. Schiffman. (1997). *No Title*. Speech Acts and Conversation. https://www.sas.upenn.edu/~haroldfs/edling/handouts/speechacts/spchax2.html

Haugh, M. (2008). Intention in pragmatics. *Intercultural Pragmatics*, *5*(2), 99–110.

Haugh, M. (2012). On understandings of intention: A response to Wedgwood. *Intercultural Pragmatics*, *9*(2), 161–194.

He, J., Fang, X., Liu, H., & Li, X. (2018). Mobile app recommendation: an involvement-enhanced approach. *Available at SSRN 3279195*.

Hellbernd, N., & Sammler, D. (2016). Prosody conveys speaker's intentions: Acoustic cues for speech act perception. *Journal of Memory and Language*, *88*, 70–86.

Hochreiter, R., & Waldhauser, C. (2014). The role of emotions in propagating brands in social networks. *ArXiv Preprint ArXiv:1409.4617*.

Hong, L., Dan, O., & Davison, B. D. (2011). Predicting popular messages in twitter. *Proceedings of the 20th International Conference Companion on World Wide Web*, 57–58.

Hou, R., Pérez-Rosas, V., Loeb, S., & Mihalcea, R. (2019). Towards automatic detection of misinformation in online medical videos. *2019 International Conference on Multimodal Interaction*, 235–243.

Hutto, C. J., & Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. *Eighth International AAAI Conference on Weblogs and Social Media*.

Ilyas, S., & Khushi, Q. (2012). Facebook status updates: A speech act analysis. *Academic Research International, 3*(2), 500–507.

Irshad, S., & Soomro, T. R. (2018). Identity theft and social media. *International Journal of Computer Science and Network Security*, *18*(1), 43–55.

Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., & Westwood, S. J. (2019). The origins and consequences of affective polarization in the United States. *Annual Review of Political Science*, *22*, 129–146.

Iyengar, S., Sood, G., & Lelkes, Y. (2012). Affect, not ideologya social identity perspective on polarization. *Public Opinion Quarterly*, *76*(3), 405–431.

Jaccard, J., Wan, C. K., & Turrisi, R. (1990). The detection and interpretation of interaction effects between continuous variables in multiple regression. *Multivariate Behavioral Research*, *25*(4), 467–478.

Jack, C. (2017). Lexicon of lies: Terms for problematic information. *Data & Society*, *3*, 22.

Jenkins, H. (2009). *Confronting the challenges of participatory culture: Media education for the 21st century*. The MIT Press.

Jia, J., Wang, B., & Gong, N. Z. (2017). Random walk based fake account detection in online social networks. *2017 47th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*, 273–284.

Jost, J. T., van der Linden, S., Panagopoulos, C., & Hardin, C. D. (2018). Ideological asymmetries in conformity, desire for shared reality, and the spread of misinformation. *Current Opinion in Psychology*, *23*, 77–83.

Karimi, H., Roy, P., Saba-Sadiya, S., & Tang, J. (2018). Multi-source multi-class fake news detection. *Proceedings of the 27th International Conference on Computational Linguistics*, 1546–1557.

Karlova, N. A., & Fisher, K. E. (2013). *A social diffusion model of misinformation and disinformation for understanding human information behaviour*.

Kass-Hout, T. A., & Alhinnawi, H. (2013). Social media in public health. *Br Med Bull*, *108*(1), 5–24.

Khan, M. L. (2017). Social media engagement: What motivates user participation and consumption on YouTube? *Computers in Human Behavior*, *66*, 236–247.

Kincaid, J. P., Fishburne Jr, R. P., Rogers, R. L., & Chissom, B. S. (1975). *Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel*.

Kirk, R. T. (2015). *The buck stops in Benghazi: A practical application of the dramatistic pentad and interpersonal deception theory*.

Knobloch-Westerwick, S., & Kleinman, S. B. (2012). Preelection selective exposure: Confirmation bias versus informational utility. *Communication Research*, *39*(2), 170–193.

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

Kucharski, A. (2016). Study epidemiology of fake news. *Nature*, *540*(7634), 525.

Kuhn, M., Johnson, K., & others. (2013). *Applied predictive modeling* (Vol. 26). Springer.

Kupavskii, A., Ostroumova, L., Umnov, A., Usachev, S., Serdyukov, P., Gusev, G., & Kustarev, A. (2012). Prediction of retweet cascade size over time. *Proceedings of the 21st ACM International Conference on Information and Knowledge Management*, 2335–2338.

Kwon, S., Cha, M., Jung, K., Chen, W., & Wang, Y. (2013). Prominent features of rumor propagation in online social media. *Proceedings - IEEE International Conference on Data Mining, ICDM*, 1103–1108. https://doi.org/10.1109/ICDM.2013.61

La Rocca, G. (2020). Possible selves of a hashtag: Moving from the theory of speech acts to cultural objects to interpret hashtags. *International Journal of Sociology and Anthropology*, *12*(1), 1–9.

Lam, T., & Hsu, C. H. C. (2006). Predicting behavioral intention of choosing a travel destination. *Tourism Management*, *27*(4), 589–599.

Lampos, V., Preo\ctiuc-Pietro, D., & Cohn, T. (2013). A user-centric model of voting intention from Social Media. *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 993–1003.

Lee, J. K., Choi, J., Kim, C., & Kim, Y. (2014). Social media, network heterogeneity, and opinion polarization. *Journal of Communication*, *64*(4), 702–722.

Leung, R., Schuckert, M., & Yeung, E. (2013). Attracting user social media engagement: A study of three budget airlines Facebook pages. In *Information and communication technologies in tourism 2013* (pp. 195–206). Springer.

Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest*, *13*(3), 106–131.

Liebman, E., Saar-Tsechansky, M., & Stone, P. (2019). The Right Music at the Right Time: Adaptive Personalized Playlists Based on Sequence Modeling. *MIS Quarterly*, *43*(3).

Linvill, D. L., & Warren, P. L. (2020). Troll factories: Manufacturing specialized disinformation on Twitter. *Political Communication*, 1–21.

Littlejohn, S. W., & Foss, K. A. (2010). *Theories of human communication*. Waveland press.

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, *5*(1), 1–167.

Liu, Yang, & Wu, Y.-F. (2018). *Early Detection of Fake News on Social Media Through Propagation Path Classification with Recurrent and Convolutional Networks*.

Liu, Yu, Wang, B., Wu, B., Shang, S., Zhang, Y., & Shi, C. (2016). Characterizing super-spreading in microblog: An epidemic-based information propagation model. *Physica A: Statistical Mechanics and Its Applications*, *463*, 202–218.

Lobo, J. M., Jiménez-Valverde, A., & Real, R. (2008). AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, *17*(2), 145–151.

Luarn, P., & Lin, H.-H. (2005). Toward an understanding of the behavioral intention to use mobile banking. *Computers in Human Behavior*, *21*(6), 873–891.

Ludwig, S., & de Ruyter, K. (2016). Decoding social media speak: developing a speech act theory research agenda. *Journal of Consumer Marketing*.

Lukito, J. (2020). Coordinating a Multi-Platform Disinformation Campaign: Internet Research Agency Activity on Three US Social Media Platforms, 2015 to 2017. *Political Communication*, *37*(2), 238–255.

Maes, M., & Bischofberger, L. (2015). Will the Personalization of Online Social Networks Foster Opinion Polarization? *Available at SSRN 2553436*.

Mark M. Bailey. (n.d.). *NRCLex Library*. https://pypi.org/project/NRCLex/

Marozzo, F., & Bessi, A. (2018). Analyzing polarization of social media users and news sites during political campaigns. *Social Network Analysis and Mining*, *8*(1), 1.

Martens, B., Aguiar, L., Gomez-Herrera, E., & Mueller-Langer, F. (2018). *The digital transformation of news media and the rise of disinformation and fake news*.

Marwick, A., & Lewis, R. (2017). Media manipulation and disinformation online. *New York: Data & Society Research Institute*.

Matakos, A., Terzi, E., & Tsaparas, P. (2017). Measuring and moderating opinion polarization in social networks. *Data Mining and Knowledge Discovery*, *31*(5), 1480–1505.

McCornack, S. A. (1992). Information manipulation theory. *Communications Monographs*, *59*(1), 1–16.

McNeill, A. R., & Briggs, P. (2014). Understanding Twitter influence in the health domain: a social-psychological contribution. *Proceedings of the 23rd International Conference on World Wide Web*, 673–678.

Melinger, A., & Levelt, W. J. M. (2004). Gesture and the communicative intention of the speaker. *Gesture*, *4*(2), 119–141.

Mellon, J., & Prosser, C. (2017). Twitter and Facebook are not representative of the general population: Political attitudes and demographics of British social media users. *Research & Politics*, *4*(3), 2053168017720008.

Metts, S. (1989). An exploratory investigation of deception in close relationships. *Journal of Social and Personal Relationships*, *6*(2), 159–179.

Michelitch, K. (2015). Does electoral competition exacerbate interethnic or interpartisan economic discrimination? Evidence from a field experiment in market price bargaining. *The American Political Science Review*, *109*(1), 43.

Midi, H., Sarkar, S. K., & Rana, S. (2010). Collinearity diagnostics of binary logistic regression model. *Journal of Interdisciplinary Mathematics*, *13*(3), 253–267.

Mislove, A., Lehmann, S., Ahn, Y.-Y., Onnela, J.-P., & Rosenquist, (James). (2011). Understanding the Demographics of Twitter Users. *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, *11*.

Mohammad, S. M., Kiritchenko, S., & Zhu, X. (2013). NRC-Canada: Building the state-of-the-art in sentiment analysis of tweets. *ArXiv Preprint ArXiv:1308.6242*.

Monti, F., Frasca, F., Eynard, D., Mannion, D., & Bronstein, M. M. (2019). Fake news detection on social media using geometric deep learning. *ArXiv Preprint ArXiv:1902.06673*.

Morales, A. J., Borondo, J., Losada, J. C., & Benito, R. M. (2015). Measuring political polarization: Twitter shows the two sides of Venezuela. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *25*(3), 33114.

Narayanan, V., Barash, V., Kelly, J., Kollanyi, B., Neudert, L.-M., & Howard, P. N. (2018). Polarization, partisanship and junk news consumption over social media in the us. *ArXiv Preprint ArXiv:1803.01845*.

Nastri, J., Pena, J., & Hancock, J. T. (2006). The construction of away messages: A speech act analysis. *Journal of Computer-Mediated Communication*, *11*(4), 1025–1045.

Neri, F., Aliprandi, C., Capeci, F., Cuadros, M., & By, T. (2012). Sentiment Analysis on Social Media. *ASONAM*, *12*, 919–926.

Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Lying words: Predicting deception from linguistic styles. *Personality and Social Psychology Bulletin*, *29*(5), 665–675.

Nguyen, L. T., Wu, P., Chan, W., Peng, W., & Zhang, Y. (2012). Predicting collective sentiment dynamics from time-series social media. *Proceedings of the First International Workshop on Issues of Sentiment Discovery and Opinion Mining*, 1–8.

Nicholson, S. P., Coe, C. M., Emory, J., & Song, A. V. (2016). The politics of beauty: The effects of partisan bias on physical attractiveness. *Political Behavior*, *38*(4), 883–898.

Oh, O., Kwon, K. H., & Rao, H. R. (2010). An Exploration of Social Media in Extreme Events: Rumor Theory and Twitter during the Haiti Earthquake 2010. *Icis*, *231*, 7332–7336.

Ojebode, A. (2012). Mobile Phone Deception in Nigeria: Deceivers' Skills, Truth Bias or Respondents' Greed? *American Journal of Human Ecology*, *1*(1), 1–9.

Olea, J. L. M., & Pflueger, C. (2013). A robust test for weak instruments. *Journal of Business \& Economic Statistics*, *31*(3), 358–369.

Oshikawa, R., Qian, J., & Wang, W. Y. (2018). A survey on natural language processing for fake news detection. *ArXiv Preprint ArXiv:1811.00770*.

Ozbay, F. A., & Alatas, B. (2020). Fake news detection within online social media using supervised artificial intelligence algorithms. *Physica A: Statistical Mechanics and Its Applications*, *540*, 123174.

Parikh, S. B., & Atrey, P. K. (2018). Media-rich fake news detection: A survey. *2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, 436–441.

Parish, E. J., & Duraisamy, K. (2016). A paradigm for data-driven predictive modeling using field inversion and machine learning. *Journal of Computational Physics*, *305*, 758–774.

Paschen, J. (2019). Investigating the emotional appeal of fake news using artificial intelligence and human contributions. *Journal of Product \& Brand Management*.

Peduzzi, P. N., Hardy, R. J., & Holford, T. R. (1980). A stepwise variable selection procedure for nonlinear regression models. *Biometrics*, 511–516.

Penco, C., & Beaney, M. (2009). *Explaining the mental: naturalist and non-naturalist approaches to mental acts and processes*. Cambridge Scholars Publishing.

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

Pennycook, G., & Rand, D. G. (2019). Fighting misinformation on social media using crowdsourced judgments of news source quality. *Proceedings of the National Academy of Sciences*, *116*(7), 2521–2526.

Pennycook, G., & Rand, D. G. (2020). Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking. *Journal of Personality*, *88*(2), 185–200.

Pérez-Rosas, V., Kleinberg, B., Lefevre, A., & Mihalcea, R. (2017). Automatic detection of fake news. *ArXiv Preprint ArXiv:1708.07104*.

Perikos, I., & Hatzilygeroudis, I. (2016). Recognizing emotions in text using ensemble of classifiers. *Engineering Applications of Artificial Intelligence*, *51*, 191–201.

Petrovic, S., Osborne, M., & Lavrenko, V. (2011). RT to Win! Predicting Message Propagation in Twitter. *ICWSM*.

Prior, M. (2013). Media and political polarization. *Annual Review of Political Science*, *16*, 101–127.

Qazvinian, V., Rosengren, E., Radev, D., & Mei, Q. (2011). Rumor has it: Identifying misinformation in microblogs. *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, 1589–1599.

Qiu, J., Wu, Q., Ding, G., Xu, Y., & Feng, S. (2016). A survey of machine learning for big data processing. *EURASIP Journal on Advances in Signal Processing*, *2016*(1), 67.

Quandt, T., Frischlich, L., Boberg, S., & Schatto-Eckrodt, T. (2019). Fake news. *The International Encyclopedia of Journalism Studies*, 1–6.

Rangel, F., Giachanou, A., Ghanem, B., & Rosso, P. (2020). Overview of the 8th Author Profiling Task at PAN 2020: Profiling Fake News Spreaders on Twitter. *CLEF*.

Recanati, F. (1986). *On defining communicative intentions*.

Reis, J. C. S., Correia, A., Murai, F., Veloso, A., & Benevenuto, F. (2019a). Supervised learning for fake news detection. *IEEE Intelligent Systems*, *34*(2), 76–81.

Reis, J. C. S., Correia, A., Murai, F., Veloso, A., & Benevenuto, F. (2019b). Explainable machine learning for fake news detection. *Proceedings of the 10th ACM Conference on Web Science*, 17–26.

Richiardi, J., Achard, S., Bunke, H., & Van De Ville, D. (2013). Machine learning with brain graphs: predictive modeling approaches for functional imaging in systems neuroscience. *IEEE Signal Processing Magazine*, *30*(3), 58–70.

Ruchansky, N., Seo, S., & Liu, Y. (2017). Csi: A hybrid deep model for fake news detection. *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, 797–806.

Sadock, J. (2004). 3 Speech Acts. *The Handbook of Pragmatics*, 53.

Sbisà, M. (2001). Illocutionary force and degrees of strength in language use. *Journal of Pragmatics*, *33*(12), 1791–1814.

Schegloff, E. A. (1988). Description in the social sciences I: Talk-in-interaction. *IPrA Papers in Pragmatics*, *2*(1–2), 1–24.

Searle, J. R. (1965). What is a speech act. *Perspectives in the Philosophy of Language: A Concise Anthology*, *2000*, 253–268.

Searle, J. R. (1976). A classification of illocutionary acts. *Language in Society*, 1–23.

Searle, J. R., & Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language* (Vol. 626). Cambridge university press.

Seifert, C. M. (2002). The continued influence of misinformation in memory: What makes a correction effective? In *Psychology of learning and motivation* (Vol. 41, pp. 265–292). Elsevier.

Selwyn, N. (2012). Social media in higher education. *The Europa World of Learning*, *1*, 1–10.

Shaheen, A. A. M., & Myers, R. P. (2007). Diagnostic accuracy of the aspartate aminotransferase-to-platelet ratio index for the prediction of hepatitis C—related fibrosis: A systematic review. *Hepatology*, *46*(3), 912–921.

Shao, C., Ciampaglia, G. L., Varol, O., Flammini, A., & Menczer, F. (2017). The spread of fake news by social bots. *ArXiv Preprint ArXiv:1707.07592*, *96*, 104.

Sharif, W., Samsudin, N. A., Deris, M. M., & Naseem, R. (2016). Effect of negation in sentiment analysis. *2016 Sixth International Conference on Innovative Computing Technology (INTECH)*, 718–723.

Shen, T. J., Cowell, R., Gupta, A., Le, T., Yadav, A., & Lee, D. (2019). How gullible are you? Predicting susceptibility to fake news. *Proceedings of the 10th ACM Conference on Web Science*, 287–288.

Shim, J., & Arkin, R. C. (2013). A taxonomy of robot deception and its benefits in HRI. *2013 IEEE International Conference on Systems, Man, and Cybernetics*, 2328–2335.

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

Shu, K., Wang, S., & Liu, H. (2017). Exploiting tri-relationship for fake news detection. *ArXiv Preprint ArXiv:1712.07709*, *8*.

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

Shu, K., Wang, S., & Liu, H. (2018). Understanding user profiles on social media for fake news detection. *2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, 430–435.

Shu, K., Zhou, X., Wang, S., Zafarani, R., & Liu, H. (2019). The role of user profiles for fake news detection. *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 436–439.

Sikder, O., Smith, R. E., Vivo, P., & Livan, G. (2020). A minimalistic model of bias, polarization and misinformation in social networks. *Scientific Reports*, *10*(1), 1–11.

Sloan, L., Morgan, J., Housley, W., Williams, M., Edwards, A., Burnap, P., & Rana, O. (2013). Knowing the Tweeters: Deriving sociologically relevant demographics from Twitter. *Sociological Research Online*, *18*(3), 74–84.

Soames, S. (1984). Linguistics and psychology. *Linguistics and Philosophy*, *7*(2), 155–179.

Sobkowicz, P., & Sobkowicz, A. (2012). Two-year study of emotion and communication patterns in a highly polarized political discussion forum. *Social Science Computer Review*, *30*(4), 448–469.

Søe, S. O. (2017). Algorithmic detection of misinformation and disinformation: Gricean perspectives. *Journal of Documentation*.

Sophie Lewis. (2020). *Facebook will now warn users who "like" fake coronavirus news*. CBS News. https://www.cbsnews.com/news/coronavirus-facebook-warn-users-fake-news/

Spohr, D. (2017). Fake news and ideological polarization: Filter bubbles and selective exposure on social media. *Business Information Review*, *34*(3), 150–160.

Stahl, B. C. (2006). On the Difference or Equality of Information, Misinformation, and Disinformation: A Critical Research Perspective. *Informing Science*, *9*.

Steven Bird, & Lingling Tan. (n.d.). *VADER Library*. https://www.nltk.org/api/nltk.sentiment.html#module-nltk.sentiment.vader

Stine, Z. K., & Agarwal, N. (2019). Characterizing the language-production dynamics of social media users. *Social Network Analysis and Mining*, *9*(1), 60.

Stoker, L., & Jennings, M. K. (1995). Life-cycle transitions and political participation: The case of marriage. *American Political Science Review*, 421–433.

Sunstein, C. R. (1999). The law of group polarization. *University of Chicago Law School, John M. Olin Law & Economics Working Paper*, *91*.

Sunstein, C. R. (2002). *Conformity and dissent*.

Tan, H., Ying Wang, E., & Zhou, B. O. (2014). When the use of positive language backfires: The joint effect of tone, readability, and investor sophistication on earnings judgments. *Journal of Accounting Research*, *52*(1), 273–302.

Te'eni, D. (2006). The language-action perspective as a basis for communication support systems. *Communications of the ACM*, *49*(5), 65–70.

Thackeray, R., Neiger, B. L., Smith, A. K., & Van Wagenen, S. B. (2012). Adoption and use of social media among public health departments. *BMC Public Health*, *12*(1), 1–6.

Thelwall, M. (2017). The Heart and soul of the web? Sentiment strength detection in the social web with SentiStrength. In *Cyberemotions* (pp. 119–134). Springer.

Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., & Kappas, A. (2010). Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology*, *61*(12), 2544–2558.

Tilwick, R. L. (1975). Student Self-Scheduling: An Unintentional Deception. *NASSP Bulletin, 59*(394), 114–117.

Tjostheim, I., & Waterworth, J. A. (2020). Predicting Personal Susceptibility to Phishing. *International Conference on Information Technology & Systems*, 564–575.

Torres, R., Gerhart, N., & Negahban, A. (2018). Combating fake news: An investigation of information verification behaviors on social networking sites. *Proceedings of the 51st Hawaii International Conference on System Sciences*.

Tucker, J. A., Guess, A., Barberá, P., Vaccari, C., Siegel, A., Sanovich, S., Stukal, D., & Nyhan, B. (2018). Social media, political polarization, and political disinformation: A review of the scientific literature. *Political Polarization, and Political Disinformation: A Review of the Scientific Literature (March 19, 2018)*.

Turner, R. E., Edgley, C., & Olmstead, G. (1975). Information control in conversations: Honesty is not always the best policy. *Kansas Journal of Sociology*, 69–89.

Twitter. (n.d.). *Twitter API*. https://developer.twitter.com/en/docs/twitter-api

Vicario, M. Del, Quattrociocchi, W., Scala, A., & Zollo, F. (2019). Polarization and fake news: Early warning of potential misinformation targets. *ACM Transactions on the Web (TWEB), 13*(2), 1–22.

Villarroel Ordenes, F., Ludwig, S., De Ruyter, K., Grewal, D., & Wetzels, M. (2017). Unveiling what is written in the stars: Analyzing explicit, implicit, and discourse patterns of sentiment in social media. *Journal of Consumer Research, 43*(6), 875–894.

Vilmer, J.-B. J., Escorcia, A., Guillaume, M., & Herrera, J. (2018). Information manipulation: A challenge for our democracies. *Policy Planning Staff (CAPS) of the Ministry for Europe and Foreign Affairs and the Institute for Strategic Research (IRSEM) of the Ministry for the Armed Forces, Paris*.

Vosoughi, S., & Roy, D. (2016). Tweet acts: A speech act classifier for twitter. *ArXiv Preprint ArXiv:1605.05156.*

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science, 359*(6380), 1146–1151.

Vraga, E. K., Tully, M., & Bode, L. (2020). Empowering users to respond to misinformation about Covid-19. *Media and Communication (Lisboa)*, *8*(2), 475–479.

Wagner, C., Mitter, S., Körner, C., & Strohmaier, M. (2012). When Social Bots Attack: Modeling Susceptibility of Users in Online Social Networks. *# MSM*, 41–48.

Wagner, M. C., & Boczkowski, P. J. (2019). The reception of fake news: The interpretations and practices that shape the consumption of perceived misinformation. *Digital Journalism*, *7*(7), 870–885.

Wang, Y., McKee, M., Torbica, A., & Stuckler, D. (2019). Systematic literature review on the spread of health-related misinformation on social media. *Social Science \& Medicine*, *240*, 112552.

Ward, D., & Hexmoor, H. (2003). Deception as a means for power among collaborative agents. *Int. WS on Collaborative Agents: Autonomous Agents for Collaborative Environments*, 61–66.

Warshaw, P. R., & Davis, F. D. (1985). Disentangling behavioral intention and behavioral expectation. *Journal of Experimental Social Psychology*, *21*(3), 213–228.

Wichmann, A. (2000). The attitudinal effects of prosody, and how they relate to emotion. *ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion*.

Willem J. M., L. (1989). *Speaking : From Intention to Articulation.* A Bradford Book. https://login.libproxy.uncg.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=1735&site=ehost-live

Wise, M., & Rodriguez, D. (2013). Detecting deceptive communication through computer-mediated technology: Applying interpersonal deception theory to texting behavior. *Communication Research Reports*, *30*(4), 342–346.

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

Wu, L., Morstatter, F., Carley, K. M., & Liu, H. (2019). Misinformation in social media: definition, manipulation, and detection. *ACM SIGKDD Explorations Newsletter*, *21*(2), 80–90.

Yang, C., Lin, K. H.-Y., & Chen, H.-H. (2009). Writer meets reader: Emotion analysis of social media from both the writer's and reader's perspectives. *2009 IEEE/WIC/ACM International*

Joint Conference on Web Intelligence and Intelligent Agent Technology, *1*, 287–290.

Yang, S., Shu, K., Wang, S., Gu, R., Wu, F., & Liu, H. (2019). Unsupervised fake news detection on social media: A generative approach. *Proceedings of the AAAI Conference on Artificial Intelligence*, *33*, 5644–5651.

Yardi, S., & Boyd, D. (2010). Dynamic debates: An analysis of group polarization over time on twitter. *Bulletin of Science, Technology & Society*, *30*(5), 316–327.

Yeung, L. N. T., Levine, T. R., & Nishiyama, K. (1999). Information manipulation theory and perceptions of deception in Hong Kong. *Communication Reports*, *12*(1), 1–11.

Yoo, S., Song, J., & Jeong, O. (2018). Social media contents based sentiment analysis and prediction system. *Expert Systems with Applications*, *105*, 102–111.

Žegarac, V., & Clark, B. (1999). Phatic interpretations and phatic communication. *Journal of Linguistics*, 321–346.

Zhang, R., Gao, D., & Li, W. (2011). What are tweeters doing: Recognizing speech acts in twitter. *Workshops at the Twenty-Fifth AAAI Conference on Artificial Intelligence*.

Zhang, R., Li, W., Gao, D., & Ouyang, Y. (2012). Automatic twitter topic summarization with speech acts. *IEEE Transactions on Audio, Speech, and Language Processing*, *21*(3), 649–658.

Zhang, W., & Ram, S. (2020). A Comprehensive Analysis of Triggers and Risk Factors for Asthma Based on Machine Learning and Large Heterogeneous Data Sources. *MIS Quarterly*, *44*(1).

Zhang, X., & Ghorbani, A. A. (2020). An overview of online fake news: Characterization, detection, and discussion. *Information Processing and Management*, *57*(2). https://doi.org/10.1016/j.ipm.2019.03.004

Zhao, H., Liu, G., Shi, C., & Wu, B. (2014). A Retweet Number Prediction Model Based on Followers' Retweet Intention and Influence. *2014 IEEE International Conference on Data Mining Workshop*, 952–959.

Zhao, Z., Zhao, J., Sano, Y., Levy, O., Takayasu, H., Takayasu, M., Li, D., Wu, J., & Havlin, S. (2020). Fake news propagates differently from real news even at early stages of spreading. *EPJ Data Science*, *9*(1), 7.

Zhou, L., Burgoon, J. K., Nunamaker, J. F., & Twitchell, D. (2004). Automating linguistics-based cues for detecting deception in text-based asynchronous computer-mediated

communications. *Group Decision and Negotiation*, *13*(1), 81–106.

Zhou, L., & Zhang, D. (2007). An ontology-supported misinformation model: Toward a digital misinformation library. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, *37*(5), 804–813.

Zhou, L., & Zhang, D. (2004). Building a misinformation Ontology. *IEEE/WIC/ACM International Conference on Web Intelligence (WI'04)*, 445–448.

Zhou, X., & Zafarani, R. (2018). Fake news: A survey of research, detection methods, and opportunities. *ArXiv Preprint ArXiv:1812.00315*.

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

Zhou, X., Zafarani, R., Shu, K., & Liu, H. (2019). Fake news: Fundamental theories, detection strategies and challenges. *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, 836–837.