Information Processing and Management 58 (2021) 102679

Contents lists available at ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/infoproman

Linguistic characteristics and the dissemination of misinformation in social media: The moderating effect of information richness

Cheng Zhou, Kai Li *, Yanhong Lu

Business School, Nankai University, 94, Weijin Road, Tianjin, 300,071, China

ARTICLE INFO

Keywords:
Misinformation dissemination
Linguistic characteristics
Information richness
Social media

ABSTRACT

In recent years, there has been a rise in incorrect information, or misinformation, being shared on social media. Such misinformation tends to be more eye-catching and misleading. Previous studies on detecting misinformation online have focused mainly on linguistic characteristics; however, the role of linguistic characteristics in misinformation dissemination has not yet been thoroughly researched. In this study, we propose a misinformation dissemination model that includes the direct effects of four novel linguistic characteristics on dissemination and the moderating effect of information richness. The model is tested by using 9,631 examples of misinformation collected from Sina-Weibo, the leading social media platform in China. The results indicate that compared with correct information, misinformation containing persuasive and uncertainty words is more likely to be disseminated than misinformation containing emotional and comparative words. Furthermore, in the case of information richness, the results indicate that when the misinformation includes images, the effects of persuasive, negative emotion, comparative, and uncertainty words are strengthened, while the inclusion of videos weakens the effects of the linguistic characteristics. Finally, the results of robustness check by using the Fakenewsnet data set are consistent with our hypothesis, indicating that the four linguistic characteristics proposed by this study are also suitable for the dissemination of misinformation in English. The robustness check further demonstrates our method with well generalization.

1. Introduction

Social media services are rapidly redefining the way information is created, distributed, and shared. Information can be freely shared on social media, allowing people to find useful information more effectively and personally by using social media than by using traditional search engines (Zhao et al., 2019). Because of its ease of use, low cost, and rapid rate (Zhang & Ghorbani, 2020), an increasing number of people are looking for and receiving information on social media (Ruokolainen & Widén, 2020). According to a report (Gottfried & Shearer, 2016), nearly two-thirds of American adults retrieve information via social media.

However, with the features of freedom, anonymity, and a lack of strict gatekeepers, people can post or repost any information on social media, regardless of whether the information is correct (Yin & Zhang, 2020). As a result, social media has become an ideal breeding ground for the spread of misinformation, such as fake news, rumors, and fake reviews (Zhang & Ghorbani, 2020). Social media makes the spread of unverified information among the public faster than verified information (Balmas, 2014). In recent years, misinformation has become an important concern for both the social media industry and academia because of its role in confusing

* Corresponding author. Business School, Nankai University, 94, Weijin Road, Tianjin, 300,071, China.
E-mail address: likai@nankai.edu.cn (K. Li).

https://doi.org/10.1016/j.ipm.2021.102679
Received 20 January 2021; Received in revised form 11 June 2021; Accepted 22 June 2021
Available online 24 July 2021

netizens with biased facts. For example, a piece of misinformation "eating garlic can prevent infection of COVID-19" was considered highly credible by many people on social media in 2020. Therefore, it is increasingly important to mine and understand the underlying mechanism of misinformation dissemination.

Existing literature covers topics such as fake news detection (Margolin et al., 2018), propagation processes of misinformation based on social network analysis (Vosoughi et al., 2018), and source detection of misinformation in social networks (Shelke & Attar, 2019). To date, there have been extensive studies on building an effective and automatic framework for online misinformation detection, such as *FactCheck.org* and *PolitiFact.com*. However, studies on misinformation dissemination are limited. Misinformation dissemination is another important issue rooted deeply in information practices (Karlova & Fisher, 2013). Misinformation is accepted by people usually via widespread dissemination. If misinformation does not gain much public attention, its threats are limited. Since social media has made the spread of misinformation quicker and easier (Dekker & Engbersen, 2014; Vosoughi et al., 2018), it is crucial to study misinformation dissemination in social media.

To address this gap, this paper aims to extract linguistic characteristics from misinformation content and examine how such characteristics influence misinformation dissemination behavior in individuals on social media. In this study, the characteristics studied are the words that may trigger misinformation dissemination behavior in different ways. We investigate the impact of four linguistic characteristics: persuasive words, comparative words, emotional words (including both positive and negative emotion words), and uncertainty words. Furthermore, Yin and Zhang (2020) reveal that the representational richness of a piece of information can affect its credibility and argument quality. In recent years, social media has enabled users to create and use multimedia content (such as text-only, text + images or text + videos) more easily. According to Chen et al. (2020a), the richness of information is defined as "the display format level of information on social media". Users on social media are more likely to be attracted by and repost visual content than plain text content (Chen et al., 2020a). Thus, this study proposes misinformation richness and explores its moderating role in the relationship between the four linguistic characteristics and misinformation dissemination.

The study is structured as follows: Section 2 presents the literature review on misinformation and its dissemination in the context of social media; in addition, the existing literature on the linguistic characteristics of misinformation is elaborated upon. Section 3 presents the hypotheses and the research model. Section 4 discusses the research method and reports the empirical results. Section 5 discusses the main findings, implications, and limitations of this study. Section 6 concludes the study.

2. Theoretical background

2.1. *Misinformation and its dissemination on social media*

According to prior research, misinformation is defined as "the factually incorrect or misleading information that is not backed up with evidence" (Bode &Vraga, 2015). Misinformation has become a central issue in various fields, such as health care (Ghenai & Mejova, 2018; Zhao et al., 2020) and politics (Riedel et al., 2017; Vraga et al., 2019). Compared to disinformation, which involves misleading information knowingly being created and shared to cause harm, misinformation involves fake information inadvertently being shared without an intent to cause harm (Wardle & Derakhshan, 2017). On social media, misinformation can be found in various fields, such as science and technology, military, entertainment, and health-related fields. In recent years, misinformation on social media has gained increasing attention because of its widespread use and harmful effects.

On social media platforms, a massive amount of information is generated because every user has an almost unlimited freedom to publish messages without disclosing their real identity (Shi et al., 2018). Undoubtedly, some of the social media posts will contain misinformation. Compared with correct information, misinformation is often oversimplified to attract attention. In addition, the phenomenon of information overload is universal on social media (Fu et al., 2020), and people who want to find useful information via social media may not have the knowledge and expertise to assess the accuracy of information and further identify informative and trustworthy information (Vogel, 2017). Thus, misinformation is widely shared on social media.

Since misinformation is quite easily available on social media, different aspects of disseminating misinformation, such as propagation patterns, processes and consequences, have received much attention. For propagation processes, epidemiology modeling methods have been suggested to study misinformation dissemination in networks (Wang et al., 2017b), such as the susceptible-infected-recovered (SIR) model, because of the similarities between the dissemination of misinformation in networks and the diffusion of biological infections (Yang, 2015). For example, Zhao et al. (2013) extend the SIR model by incorporating a "forgetting" factor, which refers to spreaders forgetting to spread misinformation, to examine the misinformation propagation process in the online blogging LiveJournal. Cho et al. (2019) propose a variant of the SIR model and assume that the transition from one state to the other is determined by the uncertainty level of a person's confidence in the information. In some cases, the spread of misinformation poses potential positive outcomes. For example, Rapoza (2017) report that online misinformation would cause stock price fluctuations. On the other hand, misinformation dissemination causes people to misunderstand the situation of health emergencies such as Ebola (Oyeyemi et al., 2014).

Researchers have attempted to determine the antecedents and motivations of individuals sharing misinformation. For example, Chen et al. (2015) conducted a survey and indicate that misinformation sharing is motivated by the perceived characteristics of information, self-expression, and socialization. They also show that women have a higher prevalence and intention of sharing. Valenzuela et al. (2019) reveal that political engagement is a key antecedent to sharing misinformation via a two-wave panel survey of online social media users. Khan & Idris (2019) indicate that individuals' belief in the reliability of information has a significant influence on sharing behavior without verification. Drawing on these studies, surveys and questionnaires are mainly used to collect subjective data and then apply the data to test the proposed model. Due to its limitations, this method can address only dissemination intention. Thus, there is a lack of research on the determinants of misinformation dissemination through the collection of objective data from social media. More importantly, how linguistic styles of misinformation affect dissemination behavior in individuals is unclear.

2.2. Misinformation and linguistic characteristics

According to Zhang and Ghorbani (2020), a piece of misinformation contains physical content (such as body text, picture or video) and nonphysical content (such as emotion, opinion or feeling). In recent years, a growing stream of literature in the information system (IS) field has examined the ability to detect misinformation using linguistic characteristics. For example, Zhou and Zhang (2007) found that misinformation creators use uncertain words or terms in their content. Luca and Zervas (2016) show that fake reviews tend to be more extreme than correct reviews by examining fake reviews on Yelp. In addition, financial markets and health care systems are critical domains that apply linguistic characteristics to detect misinformation (Purda & Skillicorn, 2015; Sicilia et al., 2018; Larcker & Zakolyukina, 2012; Zhao et al., 2020).

Because online misinformation is generated intentionally by misinformation creators (Zhao et al., 2014), most creators tend to use specific writing styles to avoid detection (Conroy et al., 2015) and to increase reposts of their content. Sommariva et al. (2018) classify Zika-related rumors into three categories based on textual analysis and examined the spread of the rumor. In fact, few studies have investigated the relationship between linguistic characteristics and misinformation dissemination. More importantly, there is a lack of research focusing on the linguistic characteristics of Chinese individuals. Thus, there is an opportunity for us to address this research gap.

According to Zhang and Ghorbani (2020), linguistic-based characteristics refer to the fundamental component, structure and semantics of natural language. They classify such characteristics into three categories: word-level, sentence-level and content-level characteristics. Drawing on their work, this study attempts to employ word-level characteristics to examine their influence on misinformation dissemination. In this study, word-level characteristics can be categorized as persuasive, emotional, comparative, and uncertainty words.

2.3. Media richness theory

Media richness theory (MRT), proposed by Daft and Lengel (1986), emphasizes the ability of communication media to facilitate understanding. Media richness is defined as the information load of the media, which aims to promote shared information. Later, Daft et al. (1987) divide media richness into four categories: communication by means of face-to-face conversation, telephones, written documents, and unprocessed documents. Advances in technology enable people to create and share multimedia content easily. For example, individuals can post content with plain text, images or videos; thus, the media richness varies from low to high (Chen et al., 2020a; Denktaş-Şakar & Sürücü, 2020). Due to word limitations on social media platforms (such as Twitter), individuals are likely to extend what they want to post by including complementary material such as special tags or symbols, URLs (Castillo et al., 2011), images or videos (Gupta et al., 2013; Jin et al., 2016). Thus, this study proposes information richness to capture the features of misinformation expression and further investigates its moderating role in the relationship between linguistic-based features and misinformation dissemination.

3. Hypothesis development

3.1. Persuasive words

The persuasiveness of linguistics leads to changes in people's attitudes, which in turn cause different behaviors (Ajzen, 1980). According to the persuasive model proposed by Hovland and Weiss (1951), the effect of persuasion is determined by not only the quality of the information but also the linguistic style. For example, an online posting that states that "Professor Bob has said that exercise is conducive to good health" could persuade many people because, generally, they trust professionals (Heesacker et al., 1983). Persuasive technologies have already been applied in various fields, such as health care (Hurling et al., 2007; Kaptein et al., 2012) and crowdfunding (Kuppuswamy & Bayus, 2018; Tirdatov, 2014), to influence people's behaviors. In this study, misinformation dissemination means persuasion, that is, misinformation creators using persuasive words or terms (such as specialist, trustworthiness, expert, proficient, drawing on a fact, and according to results from an experiment) to persuade people to repost the misinformation. The purpose of creators is to bring about social changes in attitudes and behaviors (Lockton et al., 2008). Thus, we hypothesize that

H1. *. Compared to true information, misinformation that contains more persuasive words is more likely to be disseminated.*

3.2. Emotional words

People actually prefer to express their feelings when writing articles (Tang & Chen, 2012). Misinformation creators often aim to draw wide public attention and to propagate misinformation extensively (Guo et al., 2019). To achieve this goal, creators usually post misinformation with intense emotions that trigger high arousal in the public. In recent years, an increasing number of studies have exploited emotions for online misinformation detection (Bhutani et al., 2019; Guo et al., 2019; Kula et al., 2020). For example, misinformation creators tend to use more sentiment words to evoke an emotional response. In addition, compared to correct information, misinformation contains more negative emotion words (Long et al., 2017). Ghanem et al. (2020) indicate that misinformation has different emotional patterns, and emotions play a critical role in deceiving the users of social media. Surprisingly, few studies have paid attention to the effect of emotional words or terms included in misinformation on its dissemination on social media. According to the Social Sharing of Emotion theory (Rimé, 2009). Previous studies have revealed that emotional texts facilitate individuals' social media engagement behaviors, including likes (Ji et al., 2019), shares (Tang et al., 2019) and comments (Ji et al., 2019). Based on these arguments, we hypothesize *the following:*

H2. . *Compared to true information, misinformation that contains more emotional words is more likely to be disseminated.*

3.3. Comparative words

Comparisons are one of the most substantial ways of evaluation (Jindal & Liu, 2006). On social media, in addition to expressing his or her own opinions about an issue or entity, a person can also express opinions by comparing similar issues or entities based on their prior experiences. The purpose of such a comparison is to convince others and state that their opinions are acceptable. In recent years, comparative word or sentence mining has been adopted in different fields, such as customer reviews (Varathan et al., 2017), blogs or forums (Saritha & Pateriya, 2014), and web documents (Saritha & Pateriya, 2016). Surprisingly, there is a lack of studies that examine comparative words regarding whether they affect misinformation dissemination. In this study, misinformation creators may use many comparative words to portray misinformation as reliable. They usually claim that they have experience associated with the issue or entity they are comparing. For example, an example of health misinformation about COVID-19 is "The protection of dust masks and surgical masks are same toward COVID-19 because someone has experimented". Thus, comparisons are very important in influencing individuals' attitudes and behaviors because they present precise information (Varathan et al., 2017). Accordingly, we hypothesize that

H3. . *Compared to true information, misinformation that contains more comparative words is more likely to be disseminated.*

3.4. Uncertainty words

Uncertainty exists when the details of situations are complex, unpredictable, or probabilistic (Brashers, 2001). In the context of social media, uncertainty is widespread because of the lack of face-to-face interaction among people, and such inefficiency in communication may cause cognitive dissonance in the public (Dwivedi et al., 2018). In an uncertain situation, a person may experience a lack of reliable extrinsic information to better understand the situation and thus is more likely to accept misinformation to fill the gap of knowledge. Previous studies have argued that uncertainty in the information environment contributes to misinformation spreading. For example, Oh et al. (2013) demonstrate that information uncertainty is an important feature of rumors. Liu et al. (2020) show that uncertainty words are positively related to patients' perceived usefulness of online health information. Thus, if misinformation creators use more uncertainty words or terms, such as someone, sometimes, somewhere, possible, seems to be, a person has said, and so on, in postings, others may feel that the posts are suitable for them; thus, they are more likely to repost the posts on social media. Accordingly, we hypothesize that

H4. . *Compared to true information, misinformation that contains more uncertainty words is more likely to be disseminated.*

3.5. The moderating role of information richness

In this study, information richness is defined as "the display format (text, text + images or text + videos) of misinformation". Recently, many studies have examined the relationship between the richness level of information and information dissemination behavior in the context of social media. For example, Bonsón et al. (2015) show that content posted on social media with pictures is more likely to be reposted than plain text. Ji et al. (2019) find that postings with videos increase the number of shares. Compared to plain text, visual content, such as images or videos, is examined to grasp individual attention more easily (Guidry et al., 2020; Yin & Zhang, 2020). Because multimedia content can provide supplemental information, individuals can obtain extra information to better understand the information content. According to MRT, media richness should facilitate understanding; thus, compared with text-only misinformation, content that includes images or videos tends to have fewer words and often gives people a visual cue. Thus, triggered by a high level of information richness, individuals are more likely to engage in misinformation dissemination. Based on these arguments, we hypothesize that

H5a. *The positive effect of persuasive words on misinformation dissemination will be stronger when the misinformation includes images or videos.*

H5b. . *The positive effect of emotional words on misinformation dissemination will be stronger when the misinformation includes images or videos.*

H5b. . *The positive effect of comparative words on misinformation dissemination will be stronger when the misinformation includes images or videos.*

H5d. . *The positive effect of uncertainty words on misinformation dissemination will be stronger when the misinformation includes images or* *videos.*

Our research model is presented in Fig. 1. The model includes four independent variables (persuasive words, emotional words, comparative words, and uncertainty words), one variable (information richness), and one dependent variable (misinformation dissemination). Other factors, such as the content length of misinformation, the number of fans of the misinformation creator, the number of likes of and comments on misinformation, and the number of instances in which the misinformation has mentioned others, may also affect the dissemination of misinformation on social media (Lifang et al., 2020; Shi et al., 2018). Thus, these five factors are included as control variables.

4. Research method

4.1. Data collection

This study focuses on a leading social media platform in China, namely, Sina-Weibo. The platform provides an ideal setting to investigate the misinformation dissemination process for two reasons. First, Sina-Weibo has become one of the most influential microblogs in China. By the end of 2019, there were over 510 million monthly active users. Thus, relevant information can be disseminated by users in a timely manner. Second, the platform has built a special section (named "weibopiyao" in Chinese) to gather misinformation[1], which provides data support for our research. Accordingly, the platform is suitable for our study.

Sample data in this research were collected from Sina-Weibo by crawling technology. We developed a Python-based web crawler to automatically collect the posts of the *weibo.com/weibopiyao* website. For each post, the text content; the number of reposts, likes, comments; and the number of fans of the post creator were captured. Links, images, and videos were also captured, if available, to measure information richness. In total, 9,631 posts were collected.

4.2. Operationalization of variables

Dependent variable. The dependent variable is misinformation dissemination (*Misinf_diss*). It is measured by the number of reposts of each post. Previous studies have measured individuals' information dissemination behavior using quantitative indicators such as the numbers of shares or reposts on social media (Bonsón & Ratkai, 2013; Brubaker & Wilson, 2018; Chen et al., 2020a; Khobzi et al., 2019; Shi et al., 2018). Following this logic, this measurement is suitable for this study context. Previous studies, which focus on rumor spreading through the SIR model, have indicated that when rumors have already been exposed to a wide audience, people are more likely to engage in dissemination behavior (Zhao et al., 2013). Thus, the number of views may be a critical factor that influences the reposting of each post. In this study, because Sina-Weibo does not directly provide the number of views of each post, it is difficult to crawl this measurement. Previous studies (A Rahim et al., 2019; Brubaker & Wilson, 2018; Chen et al., 2020a; Khobzi et al., 2019; Lifang et al., 2020; Shi et al., 2018; Xu & Zhang, 2018), also did not include the number of views in their research model to capture the contagiousness of a particular post because of objective difficulty in obtaining such information. In this research, we will use the number of fans of the post creator and the number of mentions the post has by other users to mitigate the bias caused by the lack of the number of views.

Independent variables. The independent variables of this study are the number of persuasive words (*Persuasive*), emotional words, comparative words (*Comparative*), and uncertainty words (*Uncertain*). In particular, previous studies have revealed that rumors or fake news show more negative emotion words (Newman et al., 2003; Rubin, 2017), but it is unclear whether positive and negative emotion words are positively related to the dissemination of misinformation on social media. This study thus classifies emotional words into two subdivisions: positive (*Positive*) and negative (*Negative*) emotion words. In this study, we did not treat emotion as a binary variable for the following reasons: first, to be in line with the other three linguistic characteristics, this treatment is consistent with Long et al. (2017); second, the misinformation creator often uses positive and negative emotion words at the same time (Long et al., 2017), and previous studies have indicated that both positive (Wang et al., 2017a), and negative emotion words increase the spread of information (Zhang & Qu, 2018). Thus, it is difficult to further explore whether positive and negative emotion words could increase the dissemination of misinformation if the emotion is treated as binary. Because there are no mature dictionaries of persuasive, emotional, comparative, and uncertainty words available at present in the Chinese context, four customized dictionaries were constructed in this study. Following the procedures of Chen et al. (2020b) and Liu et al. (2020), we performed the operationalization of the dictionaries in three rounds. In the first round, we recruited three graduate students to perform manual coding work. They were informed that the purpose of the coding was to construct the dictionaries. They highlighted relevant words or terms from misinformation in a sample of 3,000 posts in parallel. Then, we calculated the kappa statistic (Cohen, 1960), which is used to conduct the reliability test of corpus classification. The average values of four dictionaries were 0.675, 0.701, 0.731, and 0.698, which were within the substantial interval suggested by Li et al. (2018). Therefore, we constructed an initial version of the four dictionaries. In the second round, two authors marked 2,000 new posts in parallel. Similar to the first coding process, we constructed a second version of the dictionaries (the average kappa values were equal to 0.681, 0.675, 0.723, and 0.745). We then compared the second version of dictionaries with the first version and added new words or terms to the first version of dictionaries. Based on these two procedures, four dictionaries were constructed (see Table 1). In the third round, we used Jieba (Sun, 2019), a Chinese text analysis tool, to automatically extract words or terms from

[1] https://weibo.com/weibopiyao

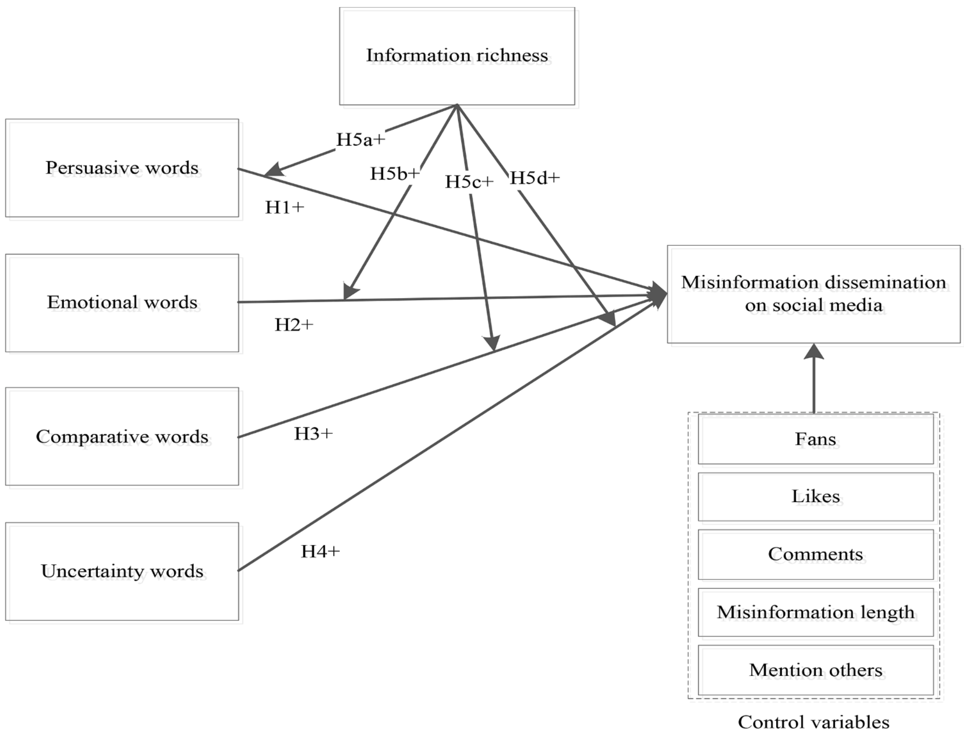

Fig. 1. The research model.

the remaining posts based on the dictionaries constructed.

Moderator variable. Prior research has classified the richness of information into three levels, namely, plain text, text + images, and text + video, which are coded as 1, 2, and 3, respectively (Chen et al., 2020a; Zhou et al., 2021). Intuitively, videos contain richer information than images, and images contain richer information than text. On social media, plain text is the basic format for posts. Videos or images are usually used to help individuals better understand the posts (Ji et al., 2019; Li et al., 2021). Therefore, we treated the information richness as categorical data and studied the two binary variables for text + images and text + video separately. For example, 1 represents a post with images or video, and 0 represents a post with plain text.

Control variables. To eliminate the interference of other factors on the results, control variables including the number of fans of the misinformation creator, the number of likes of and comments on each fake post, the length of each fake post and the number of times each fake post was mentioned by other users were set.

The definitions of the variables in this study are presented in Table 2.

4.3. *Data analysis and results*

STATA version 15.1 was used to analyze our data samples. The summary statistics of the variables are presented in Table 3. We checked for the issue of multicollinearity in STATA. As shown in Table 3, the variance inflation factor (VIF) values for all variables in this study are all below 9 (no larger than the threshold of 10), indicating that multicollinearity is not an issue in our data sample.

4.3.1. *Propensity score matching*

In this study, we aim to investigate whether misinformation containing more persuasive, emotion, comparative, and uncertainty words is more likely to be disseminated on social media. To answer this question, we first conduct propensity score matching (Caliendo

Table 1
Keyword list.

Category	Count	Examples of words in Chinese	Examples of words in English
Persuasive words	31	可信,可靠,专家,根据,事实上	Specialist, trustworthiness, credibility, according to, in fact
Emotional words	76	生气,幸运,开心,不要担心,焦虑,感激,支持	Anger, happy, luck, do not worry, concern, anxiety, appreciative, support
Comparative words	27	相似,一致,相反,相关,更好,非常,一般	Similar, consistent with, in contrary, converse, related to, associated with, better than, very, general
Uncertainty words	24		Possible, may be, seems to be, somewhere, sometime, somewhat, (what)do you think?

Table 2
Measurement of variables.

Variable name	Measure item	Description
Dependent variable		
Misinformation dissemination	*Misinf_diss*	Number of reposts of each fake post.
Independent variables		
Persuasive words	*Persuasive*	Number of persuasive words in each fake post.
Positive emotion words	*Positive*	Number of positive emotion words in each fake post.
Negative emotion words	*Negative*	Number of negative emotion words in each fake post.
Comparative words	*Comparative*	Number of comparative words in each fake post.
Uncertainty words	*Uncertainty*	Number of uncertainty words in each fake post.
Moderate variable		
Information richness	*Richness*	The richness level of each fake post.
Control variables		
Followers	*Fans*	Number of followers of the fake post creator.
Likes	*Likes*	Number of likes of each fake post.
Comments	*Comments*	Number of comments on each fake post.
Length	*Length*	Length of each fake post.
Mentions	*Mentions*	Number of times each fake post has been mentioned by other users.

Table 3
Variable statistics.

Variables	Mean	S.D.	Min	Max	VIF
Misinf_diss	316.201	3787.128	0	189,875	
Persuasive	1.252	2.313	0	26	1.22
Positive	4.116	4.067	0	10	1.81
Negative	3.731	3.563	0	13	1.79
Comparative	2.102	3.249	0	17	1.09
Uncertainty	1.008	2.479	0	7	1.43
Images	0.065	0.227	0	1	3.02
Videos	0.009	0.068	0	1	1.14
Fans	205.010	2206.104	0	106,092	8.46
Likes	198.833	2556.079	0	154,695	5.55
Comments	101.008	739.895	0	49,998	6.46
Length	67.850	56.671	10	3,715	1.06
Mentions	1.078	3.234	0	14	1.20

& Kopeining, 2008; Dehejia & Wahba, 2002) to find a matched group of true posts that cover a set of control variables similar to those targeted by fake posts. For this purpose, we randomly collected 100,000 posts (each post creator was verified by Sina-Weibo) from the recommendation section of Sina-Weibo. On Sina-Weibo, an authentication mechanism to ensure the truthfulness of posts is used. Previous studies have also indicated that the source trustworthiness of a microblogger is determined by whether user status is verified (Liu et al., 2012; Shi et al., 2018). Thus, the 100,000 posts can be seen as correct.

After constructing the sample of true posts, we performed one-on-one nearest-neighbor matching in STATA to find a matched true post for each fake post on our sample. Table 3 presents the results balance testing of the matched sample. Column (3) of Table 4 shows the biases, all of which are lower than 0.05, indicating that the matched true posts share similar characteristics to those of fake posts (D'Agostino, 1998).

4.3.2. *Hypothesis development*

Hypotheses are tested in two steps. First, the effects of linguistic characteristics on misinformation dissemination are emanated in separate regression models for fake information and correct information. Second, STATA's Seemingly Unrelated ESTimation (SUEST) postestimation command was used to combine the regression results of the separate data sets (fake and correct) and to test for

Table 4
Balance testing of matched sample.

Variable	Mean Treated	Control	Bias(%)
Images	0.053	0.039	3.0
Videos	0.035	0.035	0.0
Fans	2.895	2.835	1.6
Likes	74.959	45.231	3.4
Comments	59.562	51.620	3.9
Length	66.818	67.543	-1.5
Mentions	1.472	1.454	3.5

differences in coefficients (Weesie, 2000). As De Graaf et al. (2018) suggest, the advantage of using SUEST is that differences between groups can be examined without complicated interaction variables that increase the risk of multicollinearity between independent variables.

Based on the steps, we first conducted hierarchical regression models with three stages (Zhang et al., 2020). The following models were built to test our hypotheses:

$$\begin{aligned}&Misinf\_diss_i(Trueinf\_diss_i) = \beta_0 + \beta_1 Peruasive_i + \beta_2 Positive_i + \beta_3 Negative_i\\&+\beta_4 Comparative_i + \beta_5 Uncertain_i + \beta_6 Images_i + \beta_7 Videos_i + \beta_8 Persuasive_i * Images_i\\&+\beta_9 Persuasive_i * Videos_i + \beta_{10} Positive_i * Images_i + \beta_{11} Positive_i * Videos_i\\&+\beta_{12} Negative_i * Images_i + \beta_{13} Negative_i * Videos_i + \beta_{14} Comparative_i * Images_i\\&+\beta_{15} Comparative_i * Videos_i + \beta_{16} Uncertain_i * Images_i + \beta_{17} Uncertain_i * Videos_i + \beta' Z_i\end{aligned} \tag{1}$$

where $Misinf\_diss_i$ denotes the dissemination of misinformation; $Trueinf\_diss_i$ denotes the dissemination of correct information in this study; β parameters are the coefficients to be estimated; and Z is the vector controlling the number of followers, the number of comments on and likes of each post, the length of each post, and the number of times each post has been mentioned by other users.

The results are shown in Table 5. In stage 1 (Models 1 and 4), only the control variables were tested; in stage 2 (Models 2 and 5), the independent variables were tested; and in stage 3 (Models 3 and 6), the moderating effect was tested. Then, SUEST analyses were performed to test for differences in the coefficients between fake information and correct information. Table 6 shows the results. Integrating the results of Tables 5 and 6, we find that the effect of persuasive words ($\beta = 0.157, p < 0.001$) and uncertainty words ($\beta = 0.206, p < 0.001$) on the fake group is more significant than that on the correct group (*Persuasive:*$\beta = 0.076, p < 0.001$; *Uncertain:* $\beta = -0.032, p < 0.001$). The p-values of the t-tests are both significant (*Persuasive:*$p = 0.012$; *Uncertainty:*$p = 0.000$, see Column (5) in Table 6). This finding is consistent with our hypotheses that the more persuasive and uncertainty words included in misinformation, the more likely it will be disseminated, which indicates that H1 and H4 are supported. In terms of sentiment words, both the coefficients of *Positive* and *Negative* variables in the fake group (*Positive:*$\beta = 0.008, p > 0.05$; *Negative:* $\beta = 0.006, p < 0.05$) are smaller than in the correct group (*Positive:*$\beta = 0.016$, $p < 0.01$; *Negative:* $\beta = 0.012$, $p < 0.001$). The p-values of both t-tests are significant (*Positive:* $p = 0.04$; *Negative:* $p = 0.013$); hence, H2 is not supported. Finally, the effect of the *Comparative* variable in the fake group ($\beta = 0.109$, $p < 0.001$) is less significant than that in the correct group ($\beta = 0.121$, $p < 0.001$), and the p-value is also significant (*Comparative:* $p = 0.051$); thus, H3 is not supported.

As shown in Model 3, the interaction term *Persuasive* * *Images* ($\beta = 0.016, p < 0.001$) is positive and significant, whereas *Persuasive* * *Videos* ($\beta = -0.021, p < 0.05$) is negative and significant; hence, H5a is partially supported. For the relationship between emotion words and information richness, only the interaction term *Negative* * *Images* ($\beta = 0.019, p < 0.001$) is positive and significant, partially supporting H5b. Similarly, H5c is also partially supported because the interaction term *Comparative* * *Images* ($\beta = 0.004, p < 0.05$) is significant but *Comparative* * *Videos* ($\beta = 0.001$, $p > 0.05$) is nonsignificant. Finally, we find that the interaction term *Uncertain* * *Images* ($\beta = 0.031$, $p < 0.001$) is positive and significant, whereas *Uncertain* * *Videos* ($\beta = -0.027$, $p < 0.001$) is negative and significant; hence, H5d is partially supported.

Table 5
Regression results.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
DV:	Ln($Misinf\_diss + 1$)			Ln($Trueinf\_diss + 1$)		
Persuasive		0.157***	0.232***		0.076**	0.086***
Positive		0.021	0.018*		0.016**	-0.011
Negative		0.006*	0.005***		0.012***	0.023*
Comparative		0.109***	0.007		0.121*	0.109***
Uncertainty		0.206***	0.146***		-0.032***	-0.045***
Images			0.008***			0.009***
Videos			-0.000			-0.002*
Persuasive * *Images*			0.016***			0.065**
Persuasive * *Videos*			-0.021*			-0.087**
Positive * *Images*			0.001			0.103***
Positive * *Videos*			-0.000			0.097
Negative * *Images*			0.019***			0.023*
Negative * *Videos*			0.020			-0.030
Comparative * *Images*			0.004*			0.009
Comparative * *Videos*			0.001			0.010
Uncertainty * *Images*			0.031***			-0.056***
Uncertainty * *Videos*			-0.027**			-0.043**
Ln(*Fans*+1)	0.000***	0.000***	0.000***	0.003***	0.004***	0.005***
Ln(*Likes*+1)	-0.022**	-0.018***	-0.020***	0.013***	0.017**	0.021***
Ln(*Comments*+1)	0.003***	0.001***	0.001*	0.007***	0.003**	0.002**
Ln(*Length*)	0.009***	0.006***	0.005***	0.010***	0.008***	0.009***
Mentions	0.027***	0.015***	0.014***	0.059***	0.041***	0.035***
Observations	9,631	9,631	9,631	9,631	9,631	9,631
R-squared	0.221	0.321	0.356	0.252	0.301	0.331

Note: * $p < 0.05$; ** $p < 0.01$; ***$p < 0.001$.

Table 6
SUEST-test results.

Variable	Coefficient of fakeinformation group (β_1)	Coefficient of correctinformation group (β_2)	β_1-β_2	Difference tests (*p*-value)
Persuasive	0.157***	0.076***	0.081	0.010
Positive	0.008	0.016**	-0.008	0.004
Negative	0.006*	0.012***	-0.006	0.013
Comparative	0.019***	0.121*	-0.102	0.051
Uncertainty	0.206***	-0.032***	0.238	0.000

Note: * $p < 0.05$; ** $p < 0.01$; ***$p < 0.001$.

Furthermore, to better understand the moderating effects, the interaction diagram was plotted based on Aiken et al. (1991). Figs. 2-3 present the results. Fig. 2 shows that the dissemination of misinformation that includes images (dotted line) increases more quickly than the dissemination of misinformation that does not include images (solid line), indicating that image information richness strengthens the positive effects of persuasive, negative, comparative, and uncertainty words on misinformation dissemination. Fig. 3 indicates that there is a negative interaction between persuasive and uncertainty words and video information richness as the slope of the solid line (misinformation without videos) is larger than that of the dotted line (misinformation with videos).

4.3.3. Robustness checks

In the main analysis, we used Jieba to extract words. We conducted an initial robustness check by using an n-gram approach (Ahmed, 2017). The regression procedures were similar to the main analysis. The regression results are shown in Tables 7 and 8. Most of the effects quantitatively agree with our main findings; thus, we can be confident that the results are robust.

The main results of this study are limited to Sina-Weibo. To present the generalizability of the proposed linguistic characteristics, we conducted a second robustness check by using the Fakenewsnet data set (Shu et al., 2020). For information richness, all news articles had images in the data set; thus, we did not include the interaction term between linguistic characteristics and videos. As shown in Tables 9 and 10, the effects are mostly consistent with our main findings with the exception that the effects of positive emotion and comparative words are significant. Thus, we are more confident about our findings.

4.4. Supplementary analysis of detecting misinformation based on linguistic characteristics

As argued earlier, many studies have used linguistic characteristics to detect online misinformation (Clarke et al., 2020; Luca & Zervas, 2016; Zhang & Ghorbani, 2020). In this section, we explore whether it is feasible to detect misinformation by using the linguistic characteristics proposed by this study. As shown in Table 6, most of the characteristics are significantly different at the 5% level between misinformation and correct information. This implies that the linguistic characteristics proposed by this study can be potentially helpful in detecting misinformation on social media.

To further verify how many linguistic characteristics can contribute to the detection of misinformation, we implemented five well-known classification algorithms, including support vector machine (SVM), random forest (RF), logistic regression (LR), naive Bayes

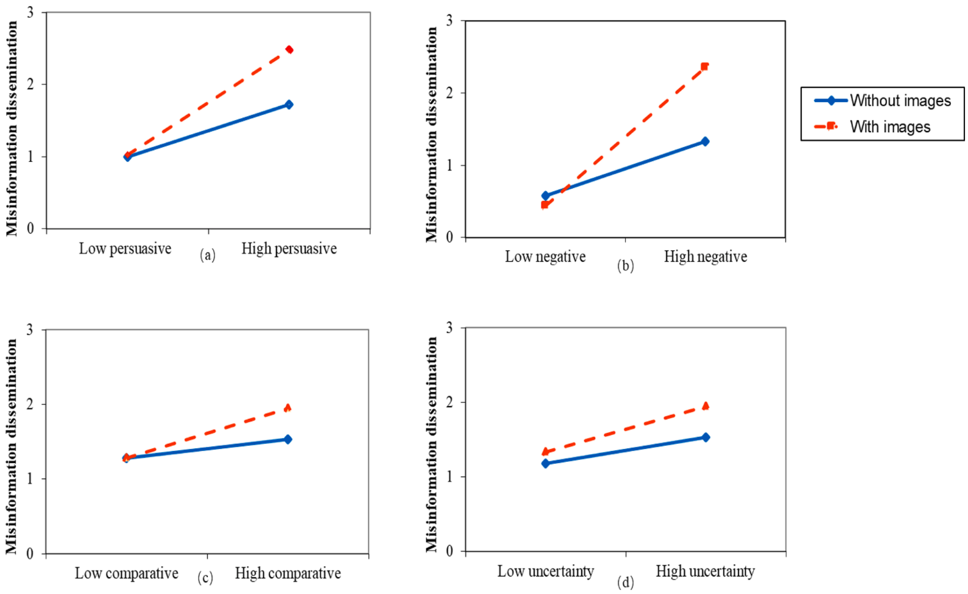

Fig. 2. The moderating effect of images on linguistic characteristics.

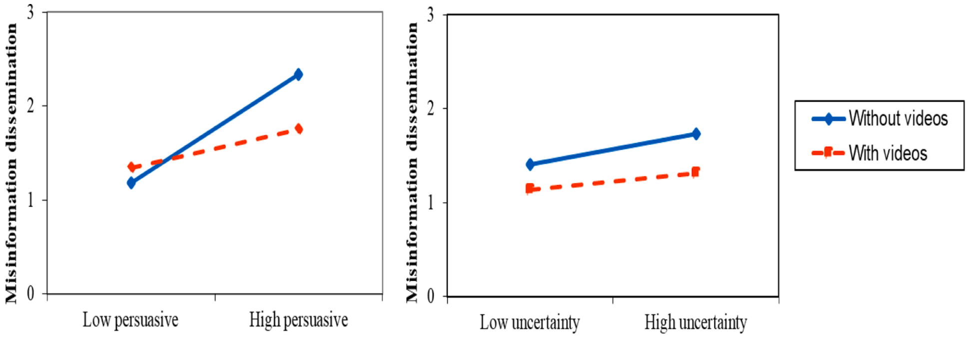

Fig. 3. The moderating effect of videos on linguistic characteristics.

Table 7
Robustness check I using n-grams.

DV:	Model 1 Model 2 Model 3 Ln(*Misinf_diss* + 1)			Model 4 Model 5 Model 6 Ln(*Trueinf_diss* + 1)		
Persuasive		0.146**	0.207**		0.106***	0.131***
Positive		-0.019	0.016*		0.016*	0.013
Negative		0.010**	0.007***		0.065**	-0.047**
Comparative		0.009***	0.010*		0.108	0.087*
Uncertainty		0.178***	0.121**		-0.125***	-0.102***
Images			0.002**			0.012**
Videos			0.000			-0.000
Persuasive Images*			0.102***			0.075***
Persuasive Videos*			-0.017**			-0.054**
Positive Images*			-0.000			0.081*
Positive Videos*			-0.000			0.102
Negative Images*			0.011*			0.001
Negative Videos*			-0.004			-0.041
Comparative Images*			0.003**			0.007*
Comparative Videos*			0.000			-0.000
Uncertainty Images*			0.113***			-0.104***
Uncertainty Videos*			-0.091**			-0.098**
Ln(*Fans*+1)	0.001**	0.003*	0.000***	0.010**	0.014***	0.025***
Ln(*Likes*+1)	0.019**	-0.021***	-0.017	0.008*	0.010***	0.009***
Ln(*Comments*+1)	0.000***	-0.001	0.002*	0.011***	0.008*	0.012
Ln(*Length*)	0.010**	0.008***	0.004*	0.011***	0.009***	0.010***
Mentions	0.016***	0.011***	0.010***	0.046***	0.039***	0.031***
Observations	9,631	9,631	9,631	9,631	9,631	9,631
R-squared	0.191	0.231	0.246	0.202	0.271	0.276

Note: * $p < 0.05$; ** $p < 0.01$; ***$p < 0.001$.

Table 8
SUEST-test results.

Variable	Coefficient of fakeinformation group (β_1)	Coefficient of correctinformation group (β_2)	β_1-β_2	Difference tests (*p*-value)
Persuasive	0.146**	0.106***	0.040	0.019
Positive	-0.019	0.016*	-0.035	0.023
Negative	0.010**	0.065**	-0.055	0.049
Comparative	0.009***	0.108	-0.099	0.062
Uncertainty	0.178***	-0.125***	0.303	0.000

Note: * $p < 0.05$; ** $p < 0.01$; ***$p < 0.001$.

(NB), and a neural network (NN), based on the six characteristics we proposed. We wrote a Python program to run both the training and testing data. Then, we split the sample of two group posts into a training data set and a testing data set at a ratio of 7:3. This ensures that both fake and true posts were equally represented in the training and testing data sets. Furthermore, we trained a classifier based on one of the five classification algorithms mentioned above on the training data set and then performed prediction on the testing data set to obtain performance measures such as precision, recall, and F1 score.

Finally, to utilize more information about legitimate news articles that are available in our sample, we repeat the previous three steps 100 times and assess the prediction performances across the 100 experiments.

Table 9
Robustness check II using Fakenewsnet data set.

DV:	Model 1 Model 2 Model 3 Ln(*Misinf_diss* + 1)			Model 4 Model 5 Model 6 Ln(*Trueinf_diss* + 1)		
Persuasive		0.127**	0.110**		0.006***	0.010***
Positive		0.013*	0.051*		0.034	0.017
Negative		0.027***	0.005***		0.054**	0.051**
Comparative		0.022**	0.006		0.100*	0.000
Uncertainty		0.141***	0.135**		-0.008***	-0.038***
Images			0.025**			0.013**
Persuasive Images*			0.027***			0.001***
Positive Images*			0.013			0.038*
Negative Images*			0.011**			0.003*
Comparative Images*			0.074**			0.008*
Uncertainty Images*			0.063***			-0.045***
Ln(*Fans*+1)	0.001**	0.000***	0.004***	0.002**	0.000***	0.000***
Ln(*Likes*+1)	0.002**	-0.001*	0.003	0.000	-0.001*	-0.000**
Ln(*Comments*+1)	0.000	0.001***	0.004**	0.000	0.001***	0.001*
Ln(*Length*)	0.011**	0.007***	0.008***	0.003***	0.006***	0.006***
Mentions	0.386***	0.408***	0.396***	0.393***	0.427***	0.396***
Observations	5,755	5,755	5,755	5,755	5,755	5,755
R-squared	0.132	0.193	0.201	0.143	0.178	0.221

Note: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Table 10
SUEST-test results.

Variable	Coefficient of fakeinformation group (β_1)	Coefficient of correctinformation group (β_2)	β_1-β_2	Difference tests (*p*-value)
Persuasive	0.127**	0.006***	0.121	0.003
Positive	0.013*	0.034	-0.021	0.009
Negative	0.027***	0.054**	-0.027	0.012
Comparative	0.022**	0.100*	-0.078	0.042
Uncertainty	0.141***	-0.008***	0.149	0.000

Note: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Table 11 summarizes the results of the 100 experiments for each of the five classification algorithms. We find that all classifiers achieved high performances (three measures are close to 70%). Further, we included the control variables of the machine learning models, and the margin of accuracy error was plus or minus 3% (see number inside the parentheses in Table 11). Although some models' accuracy was improved, such as logistic regression and naïve Bayes, the other two indicators were decreased, indicating that these models are not robust enough compared to our main models that include only the proposed linguistic characteristics. Thus, we confirm that the proposed linguistic characteristics sufficiently contribute to the detection of misinformation.

Number outside the parentheses presents the model not include the control variables.

5. Discussion

5.1. Main findings

This study yields several main findings. First, compared with correct information, misinformation with persuasive and uncertainty words is more likely to be disseminated on social media platforms. When misinformation creators write with persuasive words, individuals obtain social support and develop trust with the creators. This trust nurtures their intention to repost the misinformation. Previous studies reveal that using credible or logical words or terms (e.g., an expert has said, according to an experiment, and following this logic) can promote the effect of crowdfunding (Goering et al., 2011; Kuppuswamy & Bayu, 2018). This study extends the results of the studies and shows the role of persuasion in misinformation dissemination. In addition, with misinformation with uncertainty words, individuals cannot distinguish whether it is true or not; thus, they are likely to repost it to confirm the information's validity. This finding is consistent with the work of previous research indicating that uncertainty is a salient feature of rumor propagation (Maddock et al., 2015; Starbird et al., 2016).

Second, in terms of emotion words, we find that the relationship between negative emotion and misinformation dissemination is positive and significant. This finding is consistent with the argument of previous studies that fake news tends to use more negative emotional words (Long et al., 2017; Newman et al., 2003). However, the relationship between positive emotion and misinformation dissemination is positive but not significant. More importantly, the results of the SUR test show that compared with correct information, using more emotional words in misinformation does not enhance its dissemination. One plausible explanation is that emotional words can express the degree of compatibility and consistency between the information creators' experience and the information content (Conroy et al., 2015); however, misinformation creators prefer to use exaggerated sentiment words to attract social

Table 11
Performance of five machine learning algorithms for detecting misinformation.

Measure	Median	Mean	Minimum	Maximum
Panel A: Support vector machine				
Precision	0.683 (0.695)	0.683 (0.702)	0.671 (0.693)	0.694 (0.724)
Recall	0.683 (0.689)	0.683 (0.675)	0.672 (0.667)	0.693 (0.701)
F1	0.683 (0.699)	0.682 (0.697)	0.671 (0.689)	0.694 (0.684)
Panel B: Random forest				
Precision	0.679 (0.657)	0.679 (0.671)	0.665 (0.598)	0.690 (0.680)
Recall	0.676 (0.654)	0.676 (0.663)	0.664 (0.604)	0.686 (0.697)
F1	0.677 (0.619)	0.677 (0.620)	0.664 (0.612)	0.687 (0.629)
Panel C: Logistic regression				
Precision	0.687 (0.723)	0.688 (0.722)	0.678 (0.715)	0.702 (0.727)
Recall	0.683 (0.719)	0.684 (0.717)	0.674 (0.709)	0.698 (0.723)
F1	0.684 (0.720)	0.685 (0.718)	0.674 (0.710)	0.697 (0.724)
Panel D: Naïve Bayes				
Precision	0.685 (0.733)	0.685 (0.733)	0.673 (0.725)	0.696 (0.739)
Recall	0.612 (0.558)	0.613 (0.559)	0.602 (0.549)	0.622 (0.569)
F1	0.596 (0.500)	0.597 (0.501)	0.585 (0.486)	0.608 (0.516)
Panel E: Neural network				
Precision	0.698 (0.631)	0.698 (0.631)	0.688 (0.619)	0.707 (0.643)
Recall	0.698 (0.626)	0.699 (0.627)	0.689 (0.613)	0.709 (0.642)
F1	0.697 (0.625)	0.697 (0.626)	0.687 (0.613)	0.708 (0.642)

Note: Number inside the parentheses presents the model include the control variables.

media users' attention (Zhang & Ghorbani, 2020), and users may question anomalous emotional misinformation.

Third, we find that compared with correct information, using more comparative words in misinformation does not enhance its dissemination. This might be because misinformation creators usually have no experience with the content (Zhang & Ghorbani, 2020), so the content of comparison may be unrelated or conflicted. Such misinformation is usually perceived as incomplete or vague by users of social media (Ruokolainen & Widén, 2020); thus, they are less likely to repost it.

Fourth, our research findings suggest that information richness regarding images strengthens the effects of persuasive, negative emotional, comparative, and uncertainty words on the dissemination of misinformation. This finding indicates that compared with text-only misinformation, individuals may obtain more information from images, and their engagement with information is more easily triggered (Chen et al., 2020a). In contrast, the moderating effect of information richness regarding videos on the relationships between persuasiveness and uncertainty words and misinformation dissemination is negative. One possible explanation for this is that compared to images, seeing the content of videos requires more effort and time from individuals. In addition, according to MRT, information richness should be matched with specific contexts or tasks (Daft & Lengel, 1986; Daft et al., 1987); thus, if the content of videos is not matched with the needs of individuals, they are less likely to repost the videos on social media. These two opposing findings show the bright and dark sides of information richness and are consistent with the finding that information richness not only promotes individuals' information dissemination behavior (Lee & Xu, 2018; Shi et al., 2018) but also has a negative impact on such behavior (Chen et al., 2020a; Chung, 2017).

Finally, the results of supplementary analysis indicate that the four linguistic characteristics proposed in this study have a high performance (the accuracy of the five machine learning methods applied in this study achieves nearly 0.7) in detecting misinformation. Previous studies have verified the validity of linguistic characteristics in distinguishing misinformation (Clarke et al., 2020; Luca & Zervas, 2016; Purda & Skillicorn, 2015; Zhang & Ghorbani, 2020; Zhao et al., 2020). Our study extends the research on misinformation detection in social media by proposing new types of linguistic characteristics.

5.2. *Theoretical implications*

Several theoretical contributions of this study have been provided. First, some efforts have explored the utilization of linguistic characteristics in detecting online misinformation (Zhang & Ghorbani, 2020; Zhao et al., 2019). However, in theoretical terms, to the best of our knowledge, there has been no attempt to examine how linguistic characteristics influence misinformation dissemination. By doing so, this study proposes a model guided by a comprehensive theory to explore the underlying mechanism of misinformation in the social media context.

Second, previous studies on the linguistic characteristics of misinformation have focused on LIWC characteristics (Clarke et al., 2020), "*bag-of-words*" and "*n-grams*" (Ahmed, 2017; Conroy et al., 2015). No research has yet attempted to extract new types of characteristics in exploring misinformation dissemination on social media, especially in the Chinese context. Based on text mining, this study proposes three new types of linguistic characteristics: persuasive words, comparative words, and uncertainty words. In addition, we also include emotion words into our study. This study integrates these four characteristics to explore how they affect misinformation dissemination behavior of individuals. The findings of this study not only contribute to a deep understanding of how linguistic characteristics can affect individuals' information dissemination behavior but also highlight the need for further investigation of how specific linguistic characteristics are employed by misinformation creators on social media.

Third, this study highlights the bright and dark sides of information richness. Previous studies have indicated that individuals are easily attracted or grasped by visual content, such as images or videos, and will share such content on social media (Lee & Xu, 2018; Shi et al., 2018). However, little research has yet examined the role of richness in online misinformation and whether it can play a positive role in encouraging people to disseminate misinformation. By testing the contingent role of information richness, we find that richness positively moderates the relationship between emotional words and misinformation dissemination, while it negatively moderates the relationships between persuasive and uncertainty words and misinformation dissemination. Even though the dark side of information richness has been verified in this study, it is consistent with the finding that multimedia content can increase shares on social media (Lee & Xu, 2018; Shi et al., 2018). We also find its bright side toward limiting misinformation dissemination, which is consistent with the finding that multimedia content can also decrease shares (Chen et al., 2020a; Chung, 2017). Thus, this study enriches the information richness literature by disentangling its bright and dark sides regarding misinformation.

Finally, this study also verifies the validity of all four types of linguistic characteristics proposed in detecting misinformation. Compared with the existing literature, this study identifies that persuasive, comparative, and uncertainty words are feasible and perform well in the detection of misinformation on social media. Specifically, our findings suggest the importance of considering new types of linguistic characteristics for the detection of misinformation rather than relying on topic, sentiment, and high-dimensional textual features (e.g., "*n-grams*").

5.3. Practical implications

This study also provides several implications for practice. First, our study finds that linguistic characteristics of misinformation are positively related to misinformation dissemination. It urges users of social media to be critical when sharing postings that contain too many persuasive and uncertainty words. Further, our study also suggests that users should improve their information literacy. On social media, information literacy is described as the capability to search for information and critically evaluate it before sharing (Khan & Idris, 2019; Koltay, 2011). Thus, information-literate users evaluate online information critically.

Second, this study verifies the performance of the proposed four types of linguistic characteristics in distinguishing misinformation on social media. Thus, the proposed characteristics could be applied by social media platforms to develop and implement screening tools for misinformation.

5.4. Limitations and directions for future research

The paper has limitations that create opportunities for future research. First, future research could improve the validity of the dictionaries. The four linguistic characteristic dictionaries used in this study were manually coded, and although we tried to maintain their validity and the results of our study were accurate, some errors cannot be avoided in the coding process for the dictionaries. Second, because of the unavailability of data, we cannot completely control how many people are exposed to misinformation. Although we tried to use the number of followers and mentions that the misinformation has received from others to reduce such bias, we have to acknowledge that this method is not the best. Future research could collect relevant data from other sources to avoid this bias. Third, we tried to explore four linguistic characteristics of misinformation in the social media context. Future works could try to capture more characteristics.

6. Conclusion

In this study, we mainly focus on the relationship between the linguistic characteristics of misinformation and its dissemination on social media. In particular, we use text mining to extract four types of linguistic characteristics, persuasive words, emotional words, comparative words, and uncertainty words, and then explore their impacts on information dissemination behavior under the contingency of information richness. The results indicate that compared with correct information, the more persuasive and uncertainty words misinformation contains, the more likely it is to be disseminated. Furthermore, regarding information richness, images strengthen the effect of persuasive, negative emotional, comparative, and uncertainty words, and videos weaken the effects of persuasive and uncertainty words, indicating that there are advantages and disadvantages of information richness. Finally, all four types of characteristics can be applied in detecting misinformation. Our research method and findings contribute to the literature on misinformation dissemination and detection, linguistic characteristics of misinformation, and information richness. This study also provides significant practical implications for social media platforms, public information literacy, and the design of interventions for addressing misinformation dissemination.

CRediT authorship contribution statement

Cheng Zhou: Conceptualization, Methodology, Writing – original draft. **Kai Li:** Methodology, Writing – original draft, Supervision. **Yanhong Lu:** Data curation, Writing – review & editing.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ipm.2021.102679.

References

Rahim, A, A., I., Ibrahim, M. I., Salim, A, F., N., & Ariffin, M. A. I (2019). Health information engagement factors in Malaysia: A content analysis of Facebook use by the ministry of health in 2016 and 2017. *International Journal of Environmental Research and Public Health, 16*(4), 591.

Ahmed, H. (2017). *Detecting opinion spam and fake news using n-gram analysis and semantic similarity* (Doctoral dissertation).

Aiken, L. S., West, S. G., & Reno, R. R. (1991). *Multiple regression: Testing and interpreting interactions*. Sage.

Azjen, I. (1980). *Understanding attitudes and predicting social behavior*. Englewood Cliffs.

Balmas, M. (2014). When fake news becomes real: Combined exposure to multiple news sources and political attitudes of inefficacy, alienation, and cynicism. *Communication Research, 41*(3), 430–454.

Bhutani, B., Rastogi, N., Sehgal, P., & Purwar, A. (2019, August). Fake news detection using sentiment analysis. In *In 2019 Twelfth International Conference on Contemporary Computing (IC3)* (pp. 1–5). IEEE.

Bode, L., & Vraga, E. K. (2015). In related news, that was wrong: The correction of misinformation through related stories functionality in social media. *Journal of Communication, 65*(4), 619–638.

Bonsón, E., & Ratkai, M. (2013). A set of metrics to assess stakeholder engagement and social legitimacy on a corporate Facebook page. *Online Information Review, 37*(5), 787–803.

Bonsón, E., Royo, S., & Ratkai, M. (2015). Citizens' engagement on local governments' Facebook sites. An empirical analysis: The impact of different media and content types in Western Europe. *Government Information Quarterly, 32*(1), 52–62.

Brashers, D. E. (2001). Communication and uncertainty management. *Journal of Communication, 51*(3), 477–497.

Brubaker, P. J., & Wilson, C. (2018). Let's give them something to talk about: Global brands' use of visual content to drive engagement and build relationships. *Public Relations Review, 44*(3), 342–352.

Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys, 22*(1), 31–72.

Castillo, C., Mendoza, M., & Poblete, B. (2011, March). Information credibility on twitter. In *Proceedings of the 20th international conference on World wide web* (pp. 675–684).

Chen, Q., Min, C., Zhang, W., Wang, G., Ma, X., & Evans, R. (2020a). Unpacking the black box: How to promote citizen engagement through government social media during the COVID-19 crisis. *Computers in Human Behavior*, Article 106380.

Chen, S., Guo, X., Wu, T., & Ju, X. (2020b). Exploring the online doctor-patient interaction on patient satisfaction based on text mining and empirical analysis. *Information Processing & Management, 57*(5), Article 102253.

Chen, X., Sin, S. C. J., Theng, Y. L., & Lee, C. S. (2015). Why students share misinformation on social media: Motivation, gender, and study-level differences. *The Journal of Academic Librarianship, 41*(5), 583–592.

Cho, J. H., Rager, S., O'Donovan, J., Adali, S., & Horne, B. D. (2019). Uncertainty-based false information propagation in social networks. *ACM Transactions on Social Computing, 2*(2), 1–34.

Chung, J. E. (2017). Retweeting in health promotion: Analysis of tweets about breast cancer awareness month. *Computers in Human Behavior, 74*, 112–119.

Clarke, J., Chen, H., Du, D., & Hu, Y. J. (2020). Fake news, investor attention, and market reaction. *Information Systems Research*.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37–46.

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

Daft, R. L., & Lengel, R. H. (1986). Organizational information requirements, media richness and structural design. *Management Science, 32*(5), 554–571.

Daft, R. L., Lengel, R. H., & Trevino, L. K. (1987). Message equivocality, media selection, and manager performance: Implications for information systems. *MIS Quarterly*, 355–366.

D'Agostino, R. B., Jr. (1998). Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Statistics in Medicine, 17*(19), 2265–2281.

De Graaf, J. V. H., Hoogenboom, M., De Roos, S., & Bucx, F. (2018). Socio-demographic correlates of fathers' and mothers' parenting behaviors. *Journal of Child and Family Studies, 27*(7), 2315–2327.

Dehejia, R. H., & Wahba, S. (2002). Propensity score-matching methods for nonexperimental causal studies. *Review of Economics and Statistics, 84*(1), 151–161.

Dekker, R., & Engbersen, G. (2014). How social media transform migrant networks and facilitate migration. *Global Networks, 14*(4), 401–418.

Denktaş-Şakar, G., & Sürücü, E. (2020). Stakeholder engagement via social media: An analysis of third-party logistics companies. *The Service Industries Journal, 40*(11-12), 866–889.

Dwivedi, Y. K., Shareef, M. A., Mukerji, B., Rana, N. P., & Kapoor, K. K. (2018). Involvement in emergency supply chain for disaster management: A cognitive dissonance perspective. *International Journal of Production Research, 56*(21), 6758–6773.

Fu, S., Li, H., Liu, Y., Pirkkalainen, H., & Salo, M. (2020). Social media overload, exhaustion, and use discontinuance: Examining the effects of information overload, system feature overload, and social overload. *Information Processing & Management, 57*(6), Article 102307.

Ghanem, B., Rosso, P., & Rangel, F. (2020). An emotional analysis of false information in social media and news articles. *ACM Transactions on Internet Technology (TOIT), 20*(2), 1–18.

Ghenai, A., & Mejova, Y. (2018). Fake cures: user-centric modeling of health misinformation in social media. *Proceedings of the ACM on Human-Computer Interaction, 2*(CSCW), 1–20.

Goering, E., Connor, U. M., Nagelhout, E., & Steinberg, R. (2011). Persuasion in fundraising letters: An interdisciplinary study. *Nonprofit and Voluntary Sector Quarterly, 40*(2), 228–246.

Gottfried, J., & Shearer, E. (2016). *News use across social media platforms* 2016.

Guidry, J. P., Meganck, S. L., Lovari, A., Messner, M., Medina-Messner, V., Sherman, S., & Adams, J. (2020). Tweeting about# diseases and# publichealth: Communicating global health issues across nations. *Health Communication, 35*(9), 1137–1145.

Guo, C., Cao, J., Zhang, X., Shu, K., & Yu, M. (2019). Exploiting emotions for fake news detection on social media. *arXiv preprint arXiv:*1903.01728.

Gupta, A., Lamba, H., Kumaraguru, P., & Joshi, A. (2013, May). Faking sandy: Characterizing and identifying fake images on twitter during hurricane sandy. In *Proceedings of the 22nd international conference on World Wide Web* (pp. 729–736).

Heesacker, M., Petty, R. E., & Cacioppo, J. T. (1983). Field dependence and attitude change: Source credibility can alter persuasion by affecting message-relevant thinking. *Journal of Personality, 51*(4), 653–666.

Hovland, C. I., & Weiss, W. (1951). The influence of source credibility on communication effectiveness. *Public Opinion Quarterly, 15*(4), 635–650.

... Hurling, R., Catt, M., De Boni, M., Fairley, B., Hurst, T., Murray, P., & Sodhi, J. (2007). Using internet and mobile phone technology to deliver an automated physical activity program: Randomized controlled trial *Journal of Medical Internet Research, 9*(2), e7

Ji, Y. G., Chen, Z. F., Tao, W., & Li, Z. C. (2019). Functional and emotional traits of corporate social media message strategies: Behavioral insights from S&P 500 Facebook data. *Public Relations Review, 45*(1), 88–103.

Jin, Z., Cao, J., Zhang, Y., Zhou, J., & Tian, Q. (2016). Novel visual and statistical image features for microblogs news verification. *IEEE Transactions on Multimedia, 19*(3), 598–608.

Jindal, N., & Liu, B. (2006, August). Identifying comparative sentences in text documents. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 244–251).

Kaptein, M., De Ruyter, B., Markopoulos, P., & Aarts, E. (2012). Adaptive persuasive systems: a study of tailored persuasive text messages to reduce snacking. *ACM Transactions on Interactive Intelligent Systems (TiiS), 2*(2), 1–25.

Karlova, N. A., & Fisher, K. E. (2013). A social diffusion model of misinformation and disinformation for understanding human information behaviour. *Information Research, 18*(1), 1–12.

Khan, M. L., & Idris, I. K. (2019). Recognise misinformation and verify before sharing: A reasoned action and information literacy perspective. *Behaviour & Information Technology, 38*(12), 1194–1212.
Khobzi, H., Lau, R. Y., & Cheung, T. C. (2019). The outcome of online social interactions on Facebook pages: A study of user engagement behavior. *Internet Research, 29*(1), 2–23.
Koltay, T. (2011). The media and the literacies: Media literacy, information literacy, digital literacy. *Media, Culture & Society, 33*(2), 211–221.
Kula, S., Choraś, M., Kozik, R., Ksieniewicz, P., & Woźniak, M. (2020, June). Sentiment analysis for fake news detection by means of neural networks. In *International Conference on Computational Science* (pp. 653–666). Cham: Springer.
Kuppuswamy, V., & Bayus, B. L. (2018). Crowdfunding creative ideas: The dynamics of project backers. *The economics of crowdfunding* (pp. 151–182). Cham: Palgrave Macmillan.
Larcker, D. F., & Zakolyukina, A. A. (2012). Detecting deceptive discussions in conference calls. *Journal of Accounting Research, 50*(2), 495–540.
Lee, J., & Xu, W. (2018). The more attacks, the more retweets: Trump's and Clinton's agenda setting on Twitter. *Public Relations Review, 44*(2), 201–213.
Li, K., Zhou, C., & Yu, X. (2021). Exploring the differences of users' interaction behaviors on microblog: The moderating role of microblogger's effort. *Telematics and Informatics, 59*, Article 101553.
Li, L., Zhang, Q., Tian, J., & Wang, H. (2018). Characterizing information propagation patterns in emergencies: A case study with Yiliang Earthquake. *International Journal of Information Management, 38*(1), 34–41.
Lifang, L. I., Zhiqiang, W. A. N. G., Zhang, Q., & Hong, W. E. N. (2020). Effect of anger, anxiety, and sadness on the propagation scale of social media posts after natural disasters. *Information Processing & Management, 57*(6), Article 102313.
Liu, Y., Ren, C., Shi, D., Li, K., & Zhang, X. (2020). Evaluating the social value of online health information for third-party patients: Is uncertainty always bad? *Information Processing & Management, 57*(5), Article 102259.
Liu, Z., Liu, L., & Li, H. (2012). Determinants of information retweeting in microblogging. *Internet Research, 22*(4), 443–466.
Lockton, D., Harrison, D., & Stanton, N. (2008, June). Design with intent: Persuasive technology in a wider context. In *International Conference on Persuasive Technology* (pp. 274–278). Berlin, Heidelberg: Springer.
Long, Y., Lu, Q., Xiang, R., Li, M., & Huang, C. R. (2017, November). Fake news detection through multi-perspective speaker profiles. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers)* (pp. 252–256).
Luca, M., & Zervas, G. (2016). Fake it till you make it: Reputation, competition, and Yelp review fraud. *Management Science, 62*(12), 3412–3427.
Maddock, J., Starbird, K., Al-Hassani, H. J., Sandoval, D. E., Orand, M., & Mason, R. M. (2015, February). Characterizing online rumoring behavior using multi-dimensional signatures. In *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing* (pp. 228–241).
Margolin, D. B., Hannak, A., & Weber, I. (2018). Political fact-checking on Twitter: When do corrections have an effect? *Political Communication, 35*(2), 196–219.
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.
Oh, O., Agrawal, M., & Rao, H. R. (2013). Community intelligence and social media services: A rumor theoretic analysis of tweets during social crises. *MIS Quarterly*, 407–426.
Oyeyemi, S. O., Gabarron, E., & Wynn, R. (2014). Ebola, Twitter, and misinformation: a dangerous combination? *BMJ, 349*, g6178.
Purda, L., & Skillicorn, D. (2015). Accounting variables, deception, and a bag of words: Assessing the tools of fraud detection. *Contemporary Accounting Research, 32*(3), 1193–1223.
Rapoza, K. (2017). Can 'fake news' impact the stock market?. by Forbes. Retrieved from: https://www. forbes.com/sites/kenrapoza/2017/02/26/can-fake-newsimpact-the-stock- market/#105f09e62fac.
Riedel, B., Augenstein, I., Spithourakis, G. P., & Riedel, S. (2017). A simple but tough-to-beat baseline for the Fake News Challenge stance detection task. *arXiv preprint arXiv*:1707.03264.
Rimé, B. (2009). Emotion elicits the social sharing of emotion: Theory and empirical review. *Emotion Review, 1*(1), 60–85.
Rubin, V. L. (2017). Deception detection and rumor debunking for social media. *The SAGE handbook of social media research methods* (p. 342). Sage.
Ruokolainen, H., & Widén, G. (2020). Conceptualising misinformation in the context of asylum seekers. *Information Processing & Management, 57*(3), Article 102127.
Saritha, S. K., & Pateriya, R. K. (2014). Methods for identifying comparative sentences. *International Journal of Computer Applications, 108*(19).
Saritha, S. K., & Pateriya, R. K. (2016). Rule-based shallow parsing to identify comparative sentences from text documents. *Emerging Research in Computing, Information, Communication and Applications* (pp. 355–365). Singapore: Springer.
Shelke, S., & Attar, V. (2019). Source detection of rumor in social network–a review. *Online Social Networks and Media, 9*, 30–42.
Shi, J., Hu, P., Lai, K. K., & Chen, G. (2018). Determinants of users' information dissemination behavior on social networking sites: An elaboration likelihood model perspective. *Internet Research, 28*(2), 393–418.
Shu, K., Mahudeswaran, D., Wang, S., Lee, D., & Liu, H. (2020). Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media. *Big Data, 8*(3), 171–188.
Sicilia, R., Giudice, S. L., Pei, Y., Pechenizkiy, M., & Soda, P. (2018). Twitter rumour detection in the health domain. *Expert Systems with Applications, 110*, 33–40.
Sommariva, S., Vamos, C., Mantzarlis, A., Đào, L. U. L., & Martinez Tyson, D. (2018). Spreading the (fake) news: exploring health messages on social media and the implications for health professionals using a case study. *American Journal of Health Education, 49*(4), 246–255.
Starbird, K., Spiro, E., Edwards, I., Zhou, K., Maddock, J., & Narasimhan, S. (2016, May). Could this be true? I think so! Expressed uncertainty in online rumoring. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 360–371).
Sun, J. (2019). Jieba (Version 0.40) [Software]. Available from https://pypi.org/project/jieba/.
Tang, X., Li, S., Gu, N., & Tan, M. (2019). Exploring repost features of police-generated microblogs through topic and sentiment analysis. *Electronic Library, 37*(4), 264–273.
Tang, Y. J., & Chen, H. H. (2012, May). Mining sentiment words from microblogs for predicting writer-reader emotion transition. In *LREC* (pp. 1226–1229).
Tirdatov, I. (2014). Web-based crowd funding: Rhetoric of success. *Technical Communication, 61*(1), 3–24.
Valenzuela, S., Halpern, D., Katz, J. E., & Miranda, J. P. (2019). The paradox of participation versus misinformation: Social media, political engagement, and the spread of misinformation. *Digital Journalism, 7*(6), 802–823.
Varathan, K. D., Giachanou, A., & Crestani, F. (2017). Comparative opinion mining: a review. *Journal of the Association for Information Science and Technology, 68*(4), 811–829.
Vogel, L. (2017). Viral misinformation threatens public health. *Canadian Medical Association Journal, 189*(50), E1567. -E1567.
Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science, 359*(6380), 1146–1151.
Vraga, E. K., Kim, S. C., & Cook, J. (2019). Testing logic-based and humor-based corrections for science, health, and political misinformation on social media. *Journal of Broadcasting & Electronic Media, 63*(3), 393–414.
Wang, C., Zhou, Z., Jin, X. L., Fang, Y., & Lee, M. K. (2017a). The influence of affective cues on positive emotion in predicting instant information sharing on microblogs: Gender as a moderator. *Information Processing & Management, 53*(3), 721–734.
Wang, X., Lin, Y., Zhao, Y., Zhang, L., Liang, J., & Cai, Z. (2017b). A novel approach for inhibiting misinformation propagation in human mobile opportunistic networks. *Peer-to- Peer Networking and Applications, 10*(2), 377–394.
Wardle, C., & Derakhshan, H. (2017). Information disorder: Toward an interdisciplinary framework for research and policy making. *Council of Europe Report, 27*, 1–107.
Weesie, J. (2000). Seemlingly unrelated estimation and the cluster-adjusted sandwich estimator. *Stata Technical Bulletin, 9*(52), 34–47.
Xu, W. W., & Zhang, C. (2018). Sentiment, richness, authority, and relevance model of information sharing during social Crises—the case of# MH370 tweets. *Computers in Human Behavior, 89*, 199–206.
Yang, Y. (2015). Global stability of VEISV propagation modeling for network worm attack. *Applied Mathematical Modelling, 39*(2), 776–780.

Yin, C., & Zhang, X. (2020). Incorporating message format into user evaluation of microblog information credibility: A nonlinear perspective. *Information Processing & Management, 57*(6), Article 102345.

Zhang, H., & Qu, C. (2018). Emotional, especially negative microblogs are more popular on the web: Evidence from an fMRI study. *Brain Imaging and Behavior*, 1–11.

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

Zhang, X., Guo, F., Xu, T., & Li, Y. (2020). What motivates physicians to share free health information on online health platforms? *Information Processing & Management, 57*(2), Article 102166.

Zhao, J., Cao, N., Wen, Z., Song, Y., Lin, Y. R., & Collins, C. (2014). # fluxflow: Visual analysis of anomalous information spreading on social media. *IEEE Transactions on Visualization and Computer Graphics, 20*(12), 1773–1782.

Zhao, L., Cui, H., Qiu, X., Wang, X., & Wang, J. (2013). SIR rumor spreading model in the new media age. *Physica A: Statistical Mechanics and its Applications, 392*(4), 995–1003.

Zhao, Y., Da, J., & Yan, J. (2020). Detecting health misinformation in online health communities: Incorporating behavioral features into machine learning based approaches. *Information Processing & Management, 58*(1), Article 102390.

Zhao, Y., Zhang, J., & Wu, M. (2019). Finding users' voice on social media: An investigation of online support groups for autism-affected users on facebook. *International Journal of Environmental Research and Public Health, 16*(23), 4804.

Zhou, C., Xiu, H., Wang, Y., & Yu, X. (2021). Characterizing the dissemination of misinformation on social media in health emergencies: An empirical study based on COVID-19. *Information Processing & Management, 58*(4), Article 102554.

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.