Proceedings of the Sixteenth International AAAI Conference on Web and Social Media (ICWSM 2022)

Know It to Defeat It: Exploring Health Rumor Characteristics and Debunking Efforts on Chinese Social Media during COVID-19 Crisis

Wenjie Yang,[1] Sitong Wang,[*2†] Zhenhui Peng,[*1] Chuhan Shi,[1] Xiaojuan Ma,[1] Diyi Yang[3]

[1]The Hong Kong University of Science and Technology
[2]Columbia University
[3]Georgia Institute of Technology
{wyangbc,zpengab,cshiag}@connect.ust.hk, sw3504@columbia.edu, mxj@cse.ust.hk, diyi.yang@cc.gatech.edu

Abstract

Health-related rumors being spread online during a public crisis may pose a serious threat to people's well-being. Existing crisis informatics research lacks in-depth insights into the characteristics of *health rumors* and *the efforts to debunk them* on social media in a pandemic. To fill this gap, we conduct a comprehensive analysis of four months of rumor-related online discussion during COVID-19 on Weibo, a Chinese microblogging site. Results suggest that the dread (cause fear) type of health rumors provoked significantly more discussions and lasted longer than the wish (raise hope) type. We further explore how four kinds of social media users (i.e., government, media, organization, and individual) combat health rumors, and identify their preferred way of sharing the debunking information and the key rhetoric strategies used in the process. We examine the relationship between debunking and rumor discussions using a Granger causality approach, and show the efficacy of debunking in suppressing rumor discussions, which is time-sensitive and varies according to rumor type and debunker. Our results can provide insights into crisis informatics and risk management on social media in pandemic settings.

Introduction

During the COVID-19 pandemic, an overabundance of rumors spread broadly on social media (Brennen et al. 2020; Islam et al. 2020). In particular, *health rumors* – unverified information concerning the practice of healthcare and medicine (Cullen 2006) – can pose a major threat to public health by misleading people into action that is potentially harmful to their well-being (Ghenai 2017). Many individuals and groups, either for self-interest or altruistic reasons, participate in countering these rumors as debunkers on social media (Brennen et al. 2020). The outcome of this war between health rumors and debunkers can have a profound impact on decisions regarding health emergencies, risk management, and policy-making during a public health crisis (Gui et al. 2017b). It is hence necessary to gain an in-depth understanding of the characteristics of health rumors and the efficacy of assorted debunking efforts.

Previous research has extensively studied the rumoring and debunking process on social media in the contexts of short-term extreme events e.g., natural disasters (Oh, Kwon, and Rao 2010; Rajdev and Lee 2015) and human-induced disasters (Arif et al. 2016, 2017). Unlike those events, public health crises could last for a long time, affect a broad geographical scope, and impact a large population (Gui et al. 2018). Thus, it would require analyzing rumor and counter-rumor efforts on a greater temporal and social scale. While recent works have looked into the dissemination of misinformation and conspiracy theories during health crises like Zika virus (Ghenai and Mejova 2017; Kou et al. 2017) and COVID-19 (Brennen et al. 2020; Islam et al. 2020; Tasnim, Hossain, and Mazumder 2020), there still lacks a systematical investigation into health rumors, their distinctive characteristics, and debunking activities in those situations.

Health rumors are typically categorized into *dread* and *wish* types according to people's underlying concerns (DiFonzo et al. 2012). As the name implies, dread rumors invoke fearsome consequences (e.g., microwave use can spread cancer), while wish rumors promote favorable outcomes (e.g., vitamin can cure cancer). Past research has argued that people tend to pay more attention to dread rumors than to wish rumors (DiFonzo and Bordia 2007), as people are psychologically more interested in bad news than good news (Baumeister et al. 2001). This conclusion was confirmed by small-scale user experiments with laypeople (DiFonzo 2008; DiFonzo et al. 2012) and medical professionals (Chua and Banerjee 2018). However, the differences between these two types of health rumors on social media have not been well studied, especially in the context of public health crises. Not to mention there is insufficient understanding of debunking efforts against them.

Past research has investigated the active engagement of particular social media user groups, such as journalists (Andrews et al. 2016) and official organizations (Starbird et al. 2018), in debunking activities related to extreme events (Chen et al. 2020). However, prior studies also pointed out that these debunkers may face dilemmas in health crises. For instance, authorities' perceptions of information during health crises could be highly uncertain (Gui et al. 2017a), leading to their insufficient involvement in risk communication (Gui et al. 2017b). Questions then arise as to what kinds

[*]These authors contributed equally.
[†]This work was done while the author was at HKUST.

of social media users are actively debunking such rumors during a health crisis, what strategies they tend to apply, and how effective these counter-efforts are.

To fill the research gaps identified above, this paper aims to address the following questions: **RQ1)** What are the characteristics of online health rumors during the COVID-19 crisis, and how do these characteristics differ in terms of dread and wish rumors? **RQ2)** What kinds of social media users have contributed to countering health rumors on social media during the pandemic, and how do the efforts of these debunkers differ? **RQ3)** What are the effects of debunking, and how do the effects vary across types of health rumors and debunkers? Knowing the answers to these underexplored questions could help better tackle the infodemic (Zarocostas 2020) in the fight against COVID-19 and other public health crises. To this end, we carry out a series of visualization and quantitative analysis on health rumor discussion and debunking posts from Weibo (i.e., a Chinese microblogging site) and various Chinese fact-checking websites. Our findings reveal that dread rumors were generally more viral in nature than wish rumors, except for extreme wish rumors. We also show that debunking is useful in forecasting and suppressing rumor discussion and that the suppression effect varies across the two rumor types and across the four kinds of debunkers, namely *individuals*, *organizations*, *media*, and *government*. Based on our findings, we provide both theoretical and practical implications for better understanding of health rumors and rumor debunking. We make our annotated data publicly available for future research[1].

Related Work

Health-related misinformation and rumors circulating on social media pose a substantial threat to the public (Smailhodzic et al. 2016), influencing people's medical decisions and even threatening their lives (Wang et al. 2019). To cope with this issue and understand the nature of health rumors, previous research has extensively investigated the spread of health rumors on diverse health topics, such as the anti-vaccine movement (Nyhan et al. 2014; Dredze, Broniatowski, and Hilyard 2016) and cancer treatments (Chen, Wang, and Peng 2018; DiFonzo et al. 2012). Based on early ideas of rumor psychology, Difonzo et al. (DiFonzo et al. 2012) classified cancer rumors into wish and dread categories based on the expected consequences. They collected rumors recalled by users of online cancer communities through questionnaires and concluded that dread rumors outnumbered wish ones. Subsequent studies also found the differences between the two types of health rumors in terms of their psychological impacts on individuals (Chua and Banerjee 2017, 2018). These findings reflect the classic psychological view that "bad is stronger than good" (Cacioppo and Berntson 1994; Baumeister et al. 2001). However, these findings have not been further validated by empirical studies on social media.

In recent years, informatics in public health crises has received much scholarly attention. A health crisis can exacerbate information uncertainty (Gui et al. 2017a) and reduce trust towards authorities (Freeman et al. 2020), very likely leading to the occurrence of conspiracy theories – another type of misinformation (Bruder et al. 2013; Van Prooijen and Jostmann 2013) that is the main focus of many recent crisis informatics studies. However, the health rumors we focus on are different and only involve health knowledge, e.g., "vitamins can cure coronavirus". The subjects and scenarios of interest in our paper fill a gap in the existing literature and deepen the understanding of health rumor characteristics.

The growing danger of misinformation on social media has prompted research on rumor interventions and debunking. Previous findings about the impact of debunking on misinformation are conflicting, ranging from the "backfire effect"(Nyhan et al. 2014) to "effective" (Shin et al. 2017). The effectiveness of debunking may correlate with many factors, e.g., the source and content of the rumor (Walter et al. 2020). In the health field, several studies support the effectiveness of debunking (Ozturk, Li, and Sakamoto 2015; Pal, Chua, and Goh 2019). For example, Ozturk et al. find the presence of refutation can decrease the likelihood of health rumor sharing (Ozturk, Li, and Sakamoto 2015). However, empirical evidence demonstrating the debunking effect on health rumors across large-scale social networks is limited.

Debunking activities on social networks often take the form of "wisdom of crowds" (Tanaka, Sakamoto, and Matsuka 2013; Arif et al. 2017), especially when traditional risk communication channels may not be available in a timely manner during crisis events. The identification of information in such cases often relies on self-correction of online crowds (Arif et al. 2017). However, different types of user groups may have different levels of engagement. For example, journalists on Twitter tended to "engage earlier and correct more" for crisis rumor debunking (Starbird et al. 2018). Similarly, a study of Weibo found that influencers were more likely to post tweets that debunked COVID-19-related conspiracy theories than ordinary users (Chen et al. 2020). In this paper, we systematically compare four different kinds of debunkers, covering citizen media, official media, and third-party fact-checkers. Past studies have shown that Chinese social media users have different levels of trust in the fact-checking information from these sources (Lu et al. 2020). Our work supports this view by uncovering differences in the behavior and effectiveness of these debunkers.

Data Preparation

Sina Weibo[2] is one of the largest microblogging websites in China, where, similar to Twitter, users can post their messages, pictures and videos publicly for instant sharing, while other users can retweet, like, and comment on these posts. Weibo acts as a central hub for Chinese Internet users to access, disseminate, and receive information and news with 560 millions monthly active users in recent years (Weibo 2020). This section presents how we 1) collected Weibo posts and health rumors related to COVID-19; 2) extracted posts related to health rumors; and 3) distinguished posts with different behaviors (i.e., discussion or debunking).

[1]https://github.com/Kelaxon/COVID19-Health-Rumor

[2]https://weibo.com/

	Verified information	Examples
Indv	Unverified users, verified individuals	*Wuhan residents*, *celebrities*
Org	Enterprises, institutions, apps, schools, websites	*Dr. DingXiang*, *universities*
Gov	Government agencies	*Police departments*
Media	Media	*Xinhua News*

Table 1: Categorical schemes and examples for individuals, organizations, governments, and media.

Type	Range	Mean	Sd	Total
Wish	1 - 38247	1147.91	3744.45	312,232
Dread	1 - 51941	2291.21	6587.82	311,605

Table 2: Number of rumor-related posts per set in the wish (N=272) and dread (N=136) categories, respectively.

Data Collection

We use a dataset of COVID-19 related Weibo posts with user information from Jan 1st, 2020 to May 1st, 2020. The dataset was provided by Qingbo Big Data Technology Co Ltd[3] through monitoring posts on Weibo and then filtering out unrelated posts using a series of COVID-related keywords, e.g., pneumonia and virus. The resulting dataset contains 100,159,355 unique posts published by 26,927,690 unique users in total. Each post entry contains textual content, a timestamp, whether it is a retweet, whether it has images or a video attached, and whether the ID of the post and author can be used to indicate uniqueness.

We further categorized these unique user accounts into four types, namely *individual*, *organization*, *government*, and *media*, based on account verification information on Weibo. The platform assigns and displays a categorical code to every account to denote their revealed identities, visible to all Weibo users. We merged the categories into four types based on user status and occupation, as shown (Table 1).

Identifying Health Rumors and Related Posts

To further extract posts related to health rumors from the initial dataset, we first compiled a comprehensive list of pandemic-related rumors circulating in Chinese social media from the archives of eleven popular Chinese fact-checking websites. Ten of them are located in mainland China (six are run by commercial companies and four by governments), and one is based in Taiwan. We scrapped 5,958 fact-checking articles published from Jan 2 to April 8, 2020 from these websites. We manually inspected the topic and veracity of each article and removed any that were confirmed as not being rumors by these sites. Next, two annotators (P1 and P2) – native Chinese speakers and familiar with our work – identified health rumors from the list (Cohen's kappa $\kappa = 0.86$) and labelled their types (i.e., *wish* or *dread*) according to the definitions in (DiFonzo et al. 2012) independently ($\kappa = 0.81$). The annotators resolved the conflicts via negotiation.

[3]http://yuqing.gsdata.cn

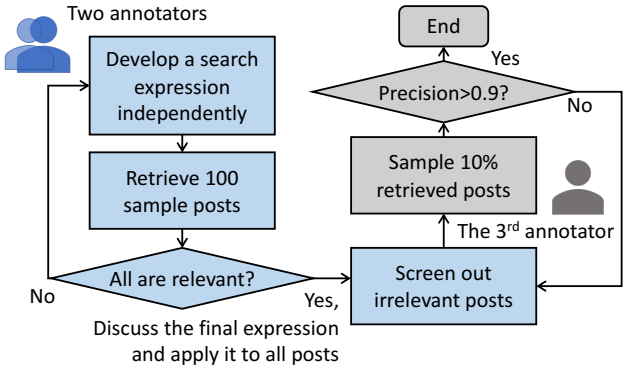

Figure 1: The workflow of identifying all posts related to a rumor topic from the Weibo dataset.

	User Count	Debunking Post Count
Individual	94,127	122,448
Government	8,675	40,225
Media	2,404	8,404
Organization	3,502	25,534

Table 3: Summary statistics for debunkers and their debunking posts.

Following the approach of (Shin et al. 2017), we use logical expressions (e.g., *Banana AND Prevent AND (Corona OR Virus)*) to retrieve rumor-related posts via Elasticsearch[4]. Figure 1 illustrates the retrieval process. For each rumor topic, P1 and P2 came up with initial expressions and manually inspected 100 random samples of the retrieved results to verify their effectiveness. They compared the expressions in the final version and further screened the results returned by these queries as follows. For Weibo posts extracted in correspondence to each of the health rumors, P1 and P2 separately scanned their contents and eliminated the irrelevant ones. A third annotator (P3) sampled 10% If the precision failed to reach 0.9 , P1 and P2 needed to repeat the screening step. We did not use recall as a quality control criterion as it was guaranteed as much as possible by the logical expression adjustment step. After dozens of rounds of iterative screening, we ended up with 408 sets of rumor-related posts (623,837 in total), each set corresponding to one health rumor topic. Table 2 shows the descriptive statistics of wish and dread posts. In the following, the term *set* or *post set* denotes a group of posts on the same topic.

Categorizing "Discuss" and "Debunk" Posts

We divided all rumor-related posts obtained in the previous step into two exclusive classes based on their behaviors: *debunk* (counter-rumor microblog messages) and *discuss* (the rest). According to our observation, debunking posts tend to explicitly declare their nature by using denials (e.g., *this is not true*) or flagging the misinformation (e.g., *this is a rumor*). We thus applied keyword matching to differentiate

[4]https://github.com/elastic/elasticsearch

Figure 2: Word clouds for health rumors, from left to right, are verbs and nouns for wish rumors and those for dread rumors; word sizes are determined by the aggregated weights in LDA.

debunk posts from *discuss* posts. Three annotators first familiarized themselves with randomly selected sets of posts (five each from wish and dread types) and collectively derived an initial list of regular expressions of debunking indicators. We then automatically marked all posts with such indicators across all the sets. After that, we used the same screening procedure mentioned earlier (10% and precision threshold 0.9) to iteratively improve the results. Eventually, we identified a total of 238,554 debunking posts in the post sets, some of which appear in more than one set. These posts were created by 108,708 distinctive debunkers (i.e., the authors of these posts, see summary statistics in Table 3). We treated all remaining posts as discussion posts, which could include affirming or neutral information about the health rumors.

Characteristics of Health Rumors (RQ1)

In RQ1, we compare the characteristics of wish and dread types of health rumors in terms of their contents and dissemination patterns.

What Concerns were Reflected in Health Rumors?

To understand what kind of public concerns are reflected in the two types of health rumors, we used Latent Dirichlet Allocation (LDA) (Blei, Ng, and Jordan 2003) to extract potential topics from rumor-related Weibo posts and visualize them by word clouds. We build different LDA models for each of the two types of rumor-related posts and then select the appropriate number of topics based on the coherence measures (Röder, Both, and Hinneburg 2015). A coherence score has no obvious improvement after increasing the number of topics to more than eight for dread rumor-related posts (score $= 0.528$), whereas the most appropriate number of topics as suggested by coherence score is 20 for wish rumor-related posts (score $= 0.510$). We observed that the top keywords under the topics identified by LDA are primarily verbs and nouns. Verbs typically indicate the types of activities people became more engrossed in during this public health crisis, and nouns convey the subjects of public attention. We thus selected the top 10 words from each topic and plotted separate verb and noun word clouds for wish and dread rumors, respectively. To stress the differences between these two types of rumors, we removed high-frequency words (i.e., "virus", "corona", "pneumonia") that are common in both rumor types and then used the aggregated weights of remaining words across their associated topics to determine their font sizes in the word clouds.

Public Concern over Rumor Topic As shown in Figure 2, from the aspect of verbs, the words with the highest weights from wish rumor-related topics included "prevention", "treatment", and "disinfect", which suggests that people are primarily interested in how to prohibit, fight against, and recover from coronavirus when discussing wish rumors. In contrast, the top words in the dread rumor-related topics include "spread", "infection", and "contact". It reflects the general public's concern about the transmission and infectivity of the virus when talking about dread rumors.

In terms of nouns in the extracted topics, "(face) mask" gained the most attention, both as a source of dread and as a source of wish, but from different perspectives. Wish rumors generally described how to "get" masks, such as DIY masks made with homemade materials, while the dread ones focused on the disposal of used masks, i.e., warning people to throw away masks used once under any circumstances. Such heated discussions around these themes may have been due to the scarcity of masks in China at the beginning of the public health crisis (SCMP 2020). The top nouns in wish rumor topics also include "antibacterial (substance)", "huanghuanglian", "hospital", "doctor", "drug", and "alcohol", showing that people may tend to gain hope from these entities. On the contrary, the top nouns contained in the dread rumor topics include "aerosol", "express delivery", "air", and "droplets", showing that the media of virus transmission often triggers fear in people.

How the Spread of Health Rumors Varied by Type?

We next compare the spread characteristics of the two kinds of health rumors. We first use a variety of single-variable linear regression models to examine the relationship between health rumors and their overall dissemination patterns. The way of using regression analysis to characterize different types of research subjects (e.g., post and user types) has been widely adopted in past work (Yang et al. 2016; Chen et al. 2020). In our case, the independent variable (IV) of each model is a categorical variable representing the type of rumors with the "wish" type as the reference group, i.e., wish $= 0$ and dread $= 1$. The dependent variables (DVs) include the following indicators. The number of observations is 408 (i.e., the total number of post sets). Following (Vosoughi, Roy, and Aral 2018), we also use curve charts to visualize the differences across rumor types in term of propagation speed.

Dependent Variables One viral nature of online misinformation is that it can be widely spread by many users over a short period of time. Following previous work (Starbird et al. 2014; Arif et al. 2016), we measured both the volume of **posts** and **users** for each health rumor by counting the post ID and the user ID. We also measured the **duration** by comparing the number of minutes between the earliest post and the last post in each post set. Since these three DVs are highly skewed, we apply a log transformation on them before analysis. Moreover, rumors threaten people's mental health by spreading panic on social networks (Ahmad and

	Post	User	Duration	Negativity
Dread	0.394***	0.397***	0.358***	0.081**
Intercept	2.099***	2.055***	2.054***	0.450***
R^2	0.039	0.040	0.037	0.024

Table 4: Prediction of dissemination patterns of health rumors. Wish type serves as the reference category. ***:$p < 0.001$; **:$p < 0.01$; *:$p < 0.05$. N=408.

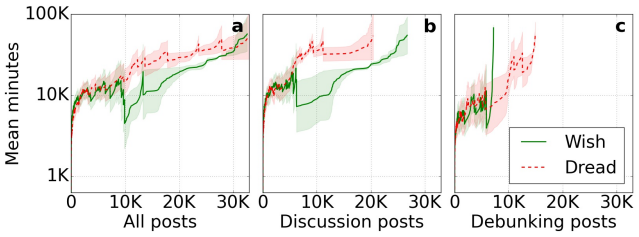

Figure 3: The average minutes it takes for wish vs dread rumors to reach any number of a) posts b) discussion posts c) debunking posts; ±1 standard errors are drawn as shadow.

Murad 2020). We quantify such **negativity** by computing the proportion of negative posts[5] in each post set.

Macro Perspective Regression analysis results (Table 4) show that, dread rumors involved significantly more posts (β = 0.394, $p < 0.001$), users (β = 0.397, $p < 0.001$) and lasted longer than wish rumors (β = 0.358, $p < 0.001$). Proportions of negative posts of dread rumors were higher than wish type (β = 0.081, $p < 0.01$).

Dynamic Perspective Figure 3 demonstrates the temporal dynamics of the spread of wish and dread rumors. These graphs illustrate the amount of time (y-axis: minutes) needed for a (wish/dread) health rumor to reach a certain level of exposure in Weibo (x-axis: number of related posts). In particular, in Fig 3a, one can see that it takes a similar length of time for wish and dread rumors to have up to 10K Weibo posts mentioning them. In fact, 96.5% of the rumors in our dataset reached fewer than 10K related posts during the four-month period. For the remaining 3.5% of extreme rumors (hitting greater than 10K posts in Fig 3a), one can clearly observe that these wish rumors spread faster than the dread ones. The green line (wish) is noticeably lower than the red line (dread), means that on average less time is spent to reach the same amount of exposure. We further split the posts based on their nature into two subsets – discussion and debunking, and measured their speed of propagation separately. Overall, the growing trend of discussion posts (Fig 3b) is similar to that in 3a with an even wider gap between dread and wish rumors. However, in Fig 3c, we found that the emergence of wish-related debunking posts slows down drastically after the post number reaches around 6.8K and nearly stops at 7K. This phenomenon comes from two "extreme" wish rumors ("Shuanghuanglian" and "alcohol can cure/prevent the coronavirus"). They are the wish rumors that received the most debunking responses in our dataset.

[5]By sentiment analysis API: http://databus.gsdata.cn/open

	Volume	Timeliness
Government	0.245***	**0.202*****
Media	**0.480*****	0.177***
Organization	0.104***	0.108***
Intercept	0.335***	3.960***
R^2	0.258	0.015

Table 5: Prediction of the engagement of debunkers. Individual serves as the reference category. ***:$p < 0.001$; **:$p < 0.01$; *:$p < 0.05$. N=108,708.

At the beginning, their debunking posts reached around 6.8K within a week, then rapidly dropped to less than ten posts a day in the following two months. Instead, debunking posts related to dread rumors, such as "takeaway food spreads the coronavirus", continued to appear in around dozens to a hundred posts per day. This indicates that while dread rumors have been continuously debunked by the public, efforts on countering wish rumors died down after some time. This might lead to wish rumors later being disseminated faster than dread rumors. We further analyze the specific effect of debunking on suppressing both types of rumors quantitatively in reply to RQ3.

To sum up, by identifying characteristics of wish and dread rumors during the pandemic (RQ1), we found differences exist between these two types of rumors in terms of their content and dynamics, which inform the rumor management agents to treat them separately according to their distinct patterns.

Identification of Debunking Efforts (RQ2)

RQ2 examines how different groups of social media users engage in dispelling health rumors during COVID-19 and the strategies they prefer to use.

How do Debunkers Engage Differently in Counter-rumor Activities on Weibo?

We built a series of single-variable linear regression models to investigate the relationship between the type of debunkers and their engagement. Model IV consists of four categories representing different debunker types, with the "individual" type as the reference. We use the following indicators as DVs. We further use descriptive statistics, such as estimated mean μ, to quantify the degree of variations. We also apply the t-test and Kolmogorov-Smirnov (KS) test to evaluate if the differences in terms of means and distributions are statistically significant. The number of observations is 108,708 (i.e., the total number of debunkers).

Dependent Variables Debunking volume and timing are essential aspects of previous studies on user debunking behavior (Starbird et al. 2018). Likewise, we analyze the **volume** and average **timeliness** of debunking posts made by the user over the four-month period. The timeliness of a debunking post is measured by the amount of time (in minutes) between it and the first post in that post set.

Engagement Results of regression models are presented in Table 5. Compared to individuals, government (β = 0.245,

	Functional Preferences					Rhetorical Preferences			
	Retweet	Duplicate	Image	Video	Length[6]	Fact	Emotion	Sarcasm	Source
Government	-1.074***	**0.743***	0.547***	**0.905***	0.095***	**3.205***	-1.920***	**-3.628***	1.092***
Media	**-3.452***	0.419***	0.718***	0.708***	**0.253***	2.687***	**-2.653***	-2.410***	**1.899***
Organization	-1.954***	-0.038	**1.075***	0.254***	0.066***	2.547***	-2.511***	-2.976**	1.079***
Intercept	0.043***	0.455***	-1.173***	-1.031***	2.128***	1.050***	-0.667***	-1.328***	0.074***
(Pseudo) R^2	0.121	0.015	0.02	0.027	0.072	0.18	0.171	0.164	0.104

Table 6: Prediction of debunkers' functional preferences (N=238,554) and rhetorical preferences (N=1,025). Individual serves as the reference category. ***:$p < 0.001$; **:$p < 0.01$; *:$p < 0.05$.

p < 0.001), media (β = 0.480, p < 0.001), and organization (β = 0.104, p < 0.001) all posted significantly more debunking posts in Weibo during the pandemic. The results of KS-test and t-test show their differences are significantly in the shape of distributions and mean (both p < 0.001), and all debunkers participated more in correcting dread rumors: Individuals (μ_{wish} = 0.55, μ_{dread} = 0.73), Government (μ_{wish} = 2.01, μ_{dread} = 2.544), Media (μ_{wish} = 4.739, μ_{dread} = 5.719), and Organization (μ_{wish} = 1.134, μ_{dread} = 1.226).

Regression results of timeliness in Table 5 show that individuals are the fastest users to publish debunking posts due to the other three having a relatively positive β. We also calculated the μ of the debunking of posts' timeliness (converted to days) from these debunkers: Individual (10.11 days), Media (12.90 days), Organization (12.73 days), and Government (13.26 days). It shows that individuals are more likely to engage in debunking activities around three days earlier than other debunkers. We further compared the timeliness of debunkers engaging in wish and dread rumors. KS test and t-test reveal that individuals and government show significant differences in timeliness between wish and dread rumors (both p<0.001). On average, individuals and governments were later to debunk dread rumors than wish rumors. Their estimated mean (convert to days) are: Individuals (μ_{wish} = 9.069, μ_{dread} = 10.966) and Government (μ_{wish} = 12.126, μ_{dread} = 14.075).

How did Posters' Preferences of Debunking Strategies Differ?

We then examine each debunking post created by users to investigate the association between their types and their preferences when editing posts, again using single-variable regression analysis with the user type as IV.

Dependent Variables The meta data of the debunking posts are used to indicate users' *functional preferences*, such as whether it is a **retweet**, whether it includes an **image** or **video**, and the **length** of the text. Users can also republish a post in a different way than through Weibo's retweet feature by copying and pasting it. Users may have different motives for such behavior, such as to have their posts appear as "original". We capture such a preference by SimHash[7] to see if a non-retweeted post is a duplicate in content to an earlier post.

For a more fine-grained description of debunking preferences, we adapt Aristotle's modes of persuasion (i.e., Logos, Pathos, Ethos) to describe how debunking messages are constructed using *rhetorical strategies* to change the attitude or behavior of others toward rumors (Petty and Cacioppo 2012). The two researchers first constructed the codebook of rhetorical strategies by a mix of inductive and deductive coding of 200 randomly sampled debunking posts. Another 1,025 posts were sampled for formal coding based on the codebook. Each post could be coded by multiple strategies. The percentage of coded strategies among the 1,025 debunking posts were: stating **facts** (88%), expressing strong **emotions** (18%), **sarcasm** (10%), and citing credible **sources** (70%); corresponding Cohen's kappa scores are 0.509, 0.699, 0.692, and 0.738. Four dichotomous variables are used to represent these strategies.

As the values of these DVs are either zero or one, we elect to use logistic regression rather than linear regression in this case. However, length is a continuous variable and we retain linear regression for it. Functional preferences are analyzed using 238,554 observations (all debunking posts), whereas rhetorical preferences are analyzed using 1025 observations (sampled debunking posts).

Functional Preferences The regression results are shown in Table 6. Compared to individuals, the other three types are significantly less likely to use the retweeting function ($\beta < 0, p < 0.001$). In addition, compared to individuals, media (β = 0.419, $p < 0.001$) and government (β = 0.743, $p < 0.001$) were significantly more likely to publish text content duplicated from some previous posts. By examining samples of duplicate posts from government debunkers, we found that they tend to quote posts from media, especially from high-impact media involving a government background, e.g., *@People's Daily*. Their quoted content can be in diverse formats, including texts, images, and videos, and are usually tagged with their original sources, for example: *"[Will shoes bring the virus home?]...no need to disinfect the soles of shoes in daily life...@People's Daily."* Combined with our previous finding that governments tended to engage in debunking activities at a later time, this may indicate the cautious attitude of governments when handling health rumors; they may like to wait until things become clearer before quoting and posting information. Compared to individuals, the other three types were more likely to use multimedia (i.e., images and video) in their debunking posts ($\beta > 0$ with

[6]Using linear regression

[7]https://github.com/yanyiwu/simhash The same hash value means having the same (or only subtle modifications) textual content with others.

	Preference	Wish μ	Wish SE	Dread μ	Dread SE	KS	t
Government	Retweet	**0.292**	0.003	0.248	0.003	0.022***	4.891***
Individual	Image	**0.359**	0.002	0.200	0.001	0.074***	30.091***
Media	Image	**0.445**	0.004	0.364	0.004	0.050***	8.068***
Individual	Video	0.176	0.001	**0.310**	0.002	0.087***	-34.175***
Individual	Length	**186.0**	1.035	166.2	0.760	0.087***	27.439***
Government	Length	**292.7**	3.203	245.7	2.380	0.098***	20.996***
Media	Length	**413.9**	4.990	314.1	3.174	0.118***	21.140***
Individual	Satire	**0.264**	0.025	0.107	0.025	0.157*	3.998***
Individual	Source	0.472	0.029	**0.610**	0.039	0.138*	-2.841**

Table 7: Statistics and hypothesis tests for function preferences (the top part) and rhetorical preferences (the bottom part) of debunking posts created by different types of debunkers relating to wish versus dread rumors. ***:$p < 0.001$; **:$p < 0.01$; *:$p < 0.05$.

individual as the reference in regression analysis for both DVs of image and video, all of $p < 0.001$). According to our observation, many debunkers chose to put arguments and explanations in pictures to avoid the character limit[8] of Weibo. They also preferred to put interviews with medical experts in the attached videos, and encourage audiences to watch the videos for more details. For example, *"Can bathing with hot water at 56°C fight the virus?.... Click on the video ↓ Experts have all the answers for you!"* Compared to individuals, all the other users preferred to write longer content in their counter-rumor posts ($\beta > 0$, p <0.001).

Between wish and dread rumors, individuals were more likely to send videos in dread-related debunking (μ_{dread} = 0.31, μ_{wish} = 0.176, $p < 0.001$ for both KS-test and t-test), and all debunkers except organizations ($p > 0.05$ for KS-test) tended to write longer texts when fighting wish rumors than they did with the dread ones ($\mu_{\text{wish}} > \mu_{\text{dread}}$, $p < 0.001$ for both tests). We present the detailed results in Table 7.

Rhetorical Preferences Table 6 also summarizes the results of regression analysis on 1,025 sampled debunking posts. Overall, the results show that government and media were significantly more likely to use facts, such as evidence and reasoning, in their debunking information compared to individuals ($\beta > 0$, $p < 0.005$). For example, *"[Can antibiotics be used to treat coronavirus?]...No, antibiotics are ineffective against viruses, only bacteria. 2019-CoV is a virus and therefore antibiotics should not be used as a means of prevention or treatment."*

Compared to individuals, the other three kinds of debunkers were significantly less likely to use strong emotion and sarcasm in their health rumor debunking posts ($\beta < 0$, $p < 0.005$). An example message from an individual debunker is: *"That shuanghuanglian is only suppress! Suppress! You have to be sick to be useful...If you get sick, the hospital will give it to you, for free!"* In contrast to individuals, all the other debunkers had a significantly higher tendency to cite credible sources ($\beta > 0$, $p < 0.005$) in their posts. We note that one typical counter-rumor strategy adopted by media was showing interviews or quotes from medical experts. The media usually specified the expert's name, affiliation, and qualification in the text. For example: *"'No evidence indicates that coronaviruses will disappear in summer', Michael Ryan, Executive Director of the WHO's Health Emergency Programme, said at a regular press conference in Geneva."* This could reflect the rigorous attitude of the media toward the sources of information.

Between wish and dread rumors, we only found significant differences in the use of sarcasm and credible source by individual debunkers in countering these two types of rumors ($p < 0.05$ for both KS-test and t-test). According to results in Table 7, individuals were significantly more likely to apply sarcasm on debunking wish rumors than dread rumors (μ_{wish} = 0.264, μ_{dread} = 0.107). For instance, *"Rumors are everywhere. Some say that the virus cannot live if you set air conditioning at 20°C, ..., I can also think of one, eating spicy chips can prevent viral infection."* This suggests that individuals may treat these wish rumors, but not so much the dread ones, in a joking manner. Besides, individuals referenced reliable sources to support their point when fighting dread rumors significantly more than when they were faced with wish rumors (μ_{wish} = 0.472, μ_{dread} = 0.61). For instance: *"WHO says there is no evidence that dogs and cats can get and spread COVID-19. People who abandon their pets, can't you read?!!!"*

To sum up, by comparing the debunking efforts of four kinds of social media users (RQ2), we found that individual debunkers seemed to engage earlier in health rumor combating and retweet the most; organizations preferred to use images to debunk rumors; media tend to contribute more and longer content; while government accounts were more likely to echo previous posts and use videos. Individuals were more likely to use emotion-based strategies such as strong emotion and satire, while other debunkers tended to use facts and cite sources. This sheds light on the use of crowd wisdom for misinformation control in social media.

Effectiveness of Debunking Efforts (RQ3)

In RQ3, we investigate the effect of debunking on health-related rumors through a combination of two analytical methods. First, we use *Granger causality analysis* (Granger 1969) to test whether the trends of debunking posts help pre-

[8]Unlike Twitter, Weibo allows posts longer than 140 characters but requires an additional click to expand the full text.

dict the trends of discussion posts. Second, we introduce a new measure, called *suppression ratio*, to quantify how declines of rumor discussion posts vary with different factors – rumor types and user roles in particular – and over time.

Granger Causality Analysis Granger causality analysis (Granger 1969) is a statistical hypothesis test that can be used to determine the relationships between two variables by checking whether or not a time series of variable X (in this case, the trend of debunking posts) is useful in predicting variable Y (the trend of discussion posts). This test is accomplished by using n number of lags (i.e., previous values) of X to model the change in Y. A p-value from the test is used to determine whether the null hypothesis that X does not help predict (i.e., Granger-cause) Y should be rejected. Since our data are grouped into different rumor topics, we tested the Granger causality between debunking and discussion on each topic separately. Nevertheless, to demonstrate the impact of rumor debunking across different topics, we follow (Dutta, Ma, and De Choudhury 2018) to report the proportion of rumors with statistically significant results (i.e., $p < 0.05$) out of the entire rumor pool.

The specific steps for testing Granger causality under each rumor topic are as follows. We first obtain a pair of time series X and Y from social media data related to a given health rumor topic by counting the number of debunking posts and discussion posts within every six hours. Since rumor propagation usually occurs within a short period, selecting a long sampling period, for instance, one day, may lead to an aggregated time series that is too short and obscures the short-term information. Similar to (De Choudhury, Kumar, and Weber 2017), we remove topics whose series contain a gap of more than one week between two consecutive steps. Following (Lütkepohl 2005), we convert the time series into a stationary series and remove the rumor topics if the resulting series fail to satisfy the Augmented Dickey-Fuller (ADF) test ($p < 0.05$) or are found to be autocorrelated by Durbin-Watson Statistic. With the remaining data, for each pair of time series that corresponds to one rumor topic, we construct a vector auto-regressive (VAR) model to determine the optimal lag n required by the Granger causality test. The VAR model is used to predict the Y series using n lags of the Y series and n lags of the X series. We select the lag value n that maximizes the likelihood of the VAR model to make the most accurate prediction as in (De Choudhury, Kumar, and Weber 2017). In our case, the best lag value indicates that the past n steps of the debunking time series contain the most useful information for predicting the discussion series. We report the distribution of these best lags based on the VAR model across rumors in Figure 4, which in a sense reflects the time span of the debunking effects.

Results of Granger Causality Analysis In total, 144 (90 pairs from wish rumors and 54 pairs from dread rumors) were obtained and analyzed using the Granger causality test. We observed significant ($p < 0.05$) Granger causation between debunking and rumor discussion in 111 health rumors (77.1%). Separately, such causation is observed in 69 wish-type rumors (76.7%) and in 42 dread-type rumors (77.8%). This result suggests that debunking and rumoring is closely

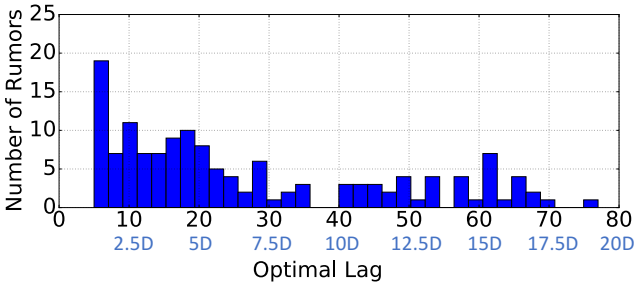

Figure 4: The histogram of the optimal lags derived from using VAR models to predict the trend of rumor discussion posts based on the trend of debunking posts. Each lag represents six hours of information. The lag value is also provided in days (D).

related, and debunking changes are useful in predicting the changes in rumor discussion. Figure 4 shows the distribution of the optimal lags from 144 constructed VAR models. In about 10% of health rumors, the best lag is 7, meaning that combining the past seven steps (i.e., 6×7 hours or 1.5 days of past information) of the debunking and discussion series can predict the subsequent discussion series the best. This is compared to combining other lengths of past information. Moreover, 60% of the best lags are ≤ 24 (6 days), and 75% of the optimal lags are ≤ 30 (7.5 days). These results suggest that using information from a relatively short previous time window (usually less than a week) for the debunking series is the most helpful in predicting changes in succeeding discussion series, possibly revealing that the association between debunking and rumor dissemination is usually short-term.

To date, our results suggest that debunking can have a short-term impact on the volume of discussions around health rumors. However, the analysis methods we have used do not fully answer the question as to whether debunking to have a suppression effect on the spread of rumors, and whether these effects vary across rumor types and debunker types. We aim to answer these questions through further suppression ratio analysis in the following subsections.

Suppression Ratio Analysis Suppression ratio is a new measure we propose to examine how effectively debunking suppresses the discussion of a rumor. We first define a post to be "*effective*" if the volume of discussion posts (i.e., non-debunking posts) on the same topic in T (a given interval) after its publication is no more than that in T before it appears on social media. This is consistent with previous ideas of measuring the decline in rumor-related posts after the appearance of debunking posts (Andrews et al. 2016; Shin et al. 2017). Given that public attention to rumors may naturally decline over time, we further calculate the ratio of effective discussion posts to all discussion posts as the baseline suppression ratio (to reflect the natural decline trend). Similarly, we calculate the suppression ratio of debunked posts. If this ratio is greater than the baseline, we consider debunking has a "*suppression effect*" on rumor-related discussion. By varying T and the source of debunked posts,

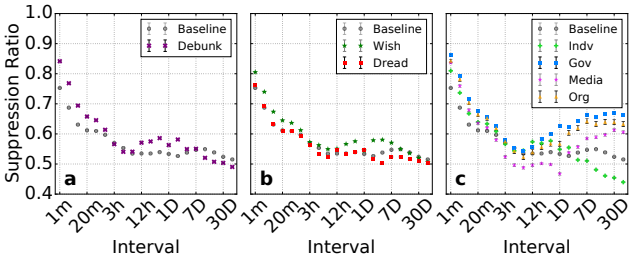

Figure 5: The suppression ratios for a) debunking posts, b) two types of health rumors, c) four kinds of debunkers, at different intervals. Error bars represent the 95% CI. The units of interval are: minutes (m), hours (h), days (D).

we can measure how debunking effectiveness is affected by these factors.

Suppression Effect In Figure 5, we define "suppression ratio" (y-axis) as the percentage of "effective" (debunking) posts and the interval (x-axis) denotes a specified length of time before and after posting for assessing suppression effectiveness. In all three subfigures, we use the suppression ratio of discussion (i.e., non-debunking) posts as the baseline, illustrated as grey dots. In general, given an interval e.g., 10 minutes, if the suppression ratio of debunking posts is above that of the baseline, we may postulate that it helps reduce the amount of discussion about health rumors to some extent.

We first investigated the effectiveness of all debunking posts as a whole in lowering health rumor discussions on different time scales. Overall, Fig 5a shows that the percentage of effective ones among all counter-rumor posts is higher than that of all non-debunking posts in the short-term ($\leq$ 1h) or medium-term (12 hours to 4 days) assessment time window. However, if we compared the amount of discussions before and after the publication of a post on the same topic in a 7+ day window, one can see that the percentage of seemingly effective debunking posts drops below that of non-debunking posts. This may indicate a tendency for a resurgence of rumor discussions for about one week after the debunking activities.

Then, we compared the effectiveness of debunking efforts on wish versus dread type of health rumors. As shown in Fig 5b, the consistently higher percentage of effort fighting wish rumors seems to result in a decrease in relevant rumor discussions than the baseline rumor-related, non-debunking posts regardless of the length of the assessment period. This may imply a positive influence of debunking activities that target wish rumors on pruning subsequent discussions. Such an observation is consistent with our finding in RQ1 that once wish rumor debunking posts cease to occur, the amount of discussions about these rumors grows at a fast rate. In contrast, the ratio of effective posts for combating dread rumors stays the same as or is lower than the baseline. In certain assessment time windows (e.g., 1h-10h, 3D-45D), a higher-than-baseline proportion (even $>$ 50% in a 3+ day window) of dread debunking posts seems to be counterproductive.

Finally, we compared rumor-suppression effectiveness across different types of rumor debunkers. As illustrated in Fig 5c, government and organization types of debunkers seem to publish a higher percentage of posts that appear to have a drop in rumor discussions within any interval length after publishing than the baseline. Their proportion of effective debunking posts first decreases as the evaluation window expands from 1 minute to 10 hours, and grows back again as the window size continues to enlarge – widening their lead over the baseline and the other two types of debunkers. Data of media debunking activities share a similar "U" shape, but with a sharper decline and then a slower come-back as the time frame increases in size. By contrast, the effect of debunking efforts by individuals seems to deteriorate when the assessment window is greater than 18 hours. In more than half of the cases, more discussions on the corresponding rumor circulate around Weibo at an interval of one week or longer after the publication of an individual-authored counter-rumor post.

In sum, we found that debunking had a short-term effect on health rumor discussions and inhibited the growth of such discussions. Debunking efforts by governments and organizations were the most persistent and effective.

Discussion

This paper examines how people discuss health rumors and engage in debunking activities on a Chinese social media during the COVID-19 health crisis. Our key findings suggest that: 1) dread rumors generally spread more virally and receive more attention than wish rumors (except for the most extreme ones); 2) when fighting rumors, individual citizens tend to be more emotional while institutions are more critical; and 3) debunking efforts are effective overall, but may sometimes have the opposite effect (e.g., counter-rumor posts by media may increase rumor discussions in the short term).

Implications

Theoretical implication for crisis informatics This study provides preliminary large-scale empirical evidence on how various kinds of online health rumors spread and are corrected in the context of health crises. First, our findings generally support the fact that dread rumors outnumber wish rumors (DiFonzo and Bordia 2007) and social media users pay more attention to them than to wish rumors (DiFonzo 2008; Chua and Banerjee 2018) under the pandemic setting. The reasons behind the differences are often explained by the negativity bias theory (Cacioppo and Berntson 1994), which suggests that people tend to attach more importance to negative news than to good news. However, our other findings on extreme rumors complicate these views – in some cases, rumors of hope can attract far more public attention. Future research is needed to investigate whether this is caused by the particular nature of a crisis. Second, we confirm the active role of specific kinds of users in rumor correction, such as journalists (Andrews et al. 2016), mainstream media, and official accounts (Starbird et al. 2018). Our analysis on rhetorical strategies proposed in persuasion theory (Petty and Cacioppo 2012) provides interesting insights into the different

debunking mechanisms applied by various social media user groups, e.g., the tendency of individuals to use sarcasm. Finally, as previous works have indicated, the effectiveness of rumor correction is time-sensitive (Lewandowsky et al. 2012; Ozturk, Li, and Sakamoto 2015). The introduction of a temporal scale of debunking effects could provide a more nuanced perspective on how rumors are suppressed by assorted debunking efforts.

Practical implication for risk management Our findings generally support the suppression effect of debunking activities on the spread of health rumors but also reveal some challenges faced in the process. On the one hand, health information seekers use social media to gain knowledge (e.g., treatments and consequence (Gui et al. 2017b)) and make health decisions (Zhang et al. 2017). However, the high degree of uncertainty associated with a health crise (Gui et al. 2017a) and the lay public's lack of expertise make it hard for them to evaluate the quality and veracity of online health information on their own (Ortutay and Klepper 2020). In such cases, tips or criticisms from credible platforms and knowledgeable users are valuable signals of information accuracy. On the other hand, choosing how to debunk rumors can be costly for risk managers (Miller 2020). Under unknown situations, the boundaries of information veracity are blurred, e.g., whether a homemade face-mask is helpful may depend on the materials. Simply criticising all such information in the face of extreme scarcity of masks may increase public anxiety. Yet, the lack of debunking can in turn, as found in our data, result in the continued spread of rumors.

For such a dilemma, more care is needed about when and how to perform risk communication in a pandemic. First, as implied in our findings, health authorities and official agencies should establish a good understanding of health-related concerns of the general public from past crisis experiences to address their information needs in a timely manner (Gui et al. 2017b). Second, while collective debunking activities on social media are seemingly effective, some users' social network structures are rather closed and exhibit homophily (McPherson, Smith-Lovin, and Cook 2001). Thus useful information from debunkers may never get circulated to them. Social media platforms may consider strategies to break rumor echo chambers, for example, by automatically attaching counter-rumor posts to previous rumor discussions (Ozturk, Li, and Sakamoto 2015). Third, even though people have a tendency to focus on "bad news" in a public health crisis, one should not overlook the harm of some inaccurate or false hope. It is important to maintain adequate debunking efforts for both kinds of misinformation. Besides, some risky ways of debunking should be avoided, such as "post first, check later" (Bruno 2011), because they may have the opposite effect and trigger more discussions about the target rumor, as shown in our data.

Generalization Some of our findings can be applied to other crisis situations as well as to other social media platforms. In terms of different popularity and spread characteristics across rumor types, dread rumors were also observed more frequently in other health crises such as the Ebola outbreaks (Allgaier and Svalastog 2015; Sell, Hosangadi, and Trotochaud 2020). Similarly, misinformation that provokes fearful emotions have also been found to spread more widely across various rumor topics, e.g., politics, financial information, and natural disasters (Vosoughi, Roy, and Aral 2018), compared to those evoking positive emotions such as trust and joy. Additionally, the active engagement of citizens and government agencies in rumor correction activities is also reported in other social media platforms such as Twitter (Andrews et al. 2016; Starbird et al. 2018). However, our findings on debunking effectiveness may be influenced by certain factors specific to the Chinese Internet context, such as people's tendency to trust the government and public institutions (Lu et al. 2020), and thus such results may not be generalized to other global social media. In spite of that, our proposed suppression ratio method for assessing debunking effects is based on the digital trajectories of rumoring and debunking campaigns and can be readily adapted to other data-driven research on crisis informatics.

Limitations and Future Work

This paper has several limitations. First, we used keyword-matching and regular expressions to identify COVID-19 related health rumor discussions and debunking posts, which may inevitably miss capturing some data due to complex, noisy usage of language on social media (Shin et al. 2017). Second, consistent with similar work (Liao and Shi 2013), we classify user accounts into four roles based on the verification information on Weibo. We did not consider more fine-grained user categories. As a preliminary study, we does not distinguish the attitude (pro-rumor or neutral) in rumor discussion posts either. Third, the analyses in this study are retrospective and correlational, and therefore cannot determine causation. Fourth, some of our regression models have low R-squared values, which can be related to the complexity of our research context. Crisis communication in the early pandemic was complicated by many factors, such as information uncertainty and low transparency on Chinese social media. This made it hard to predict rumor propagation and people's debunking behaviors. To reach reliable conclusions, we performed KS- and t-tests and found consistent results of different associations across rumor/user categories. Fifth, we measure the effectiveness of debunking based on the decline in the number of discussion posts. It is possible that these changes are caused by other factors such as the outbreak of new events diverting public attention (Andrews et al. 2016; Starbird et al. 2018). However, we assume that our data volume is large enough to show the dominant relationship between the appearance of debunking and the decline in discussion without being susceptible to other subtle factors. In the future, we will conduct controlled experiments and interviews with social media users to understand how they perceive counter-rumor information and how it changes their behavior and attitudes.

Conclusion

In this paper, we integrate quantitative and visual analyses to examine online discussions and debunking activities regarding health rumors on Weibo during the first four months

of the COVID-19 crisis. Our study found that two types of health rumors (i.e., dread and wish rumors) differed significantly in content and dissemination. We also distinguished between the different behaviors and impacts of four kinds of social media users, from ordinary individuals to government agencies, in combating health rumors. We demonstrated the effectiveness of debunking. Our findings contribute to a better understanding of how online health rumors spread and how they are influenced by debunking efforts in public health crises.

Acknowledgments

We are grateful to the anonymous reviewers for their insightful suggestions. We thank Qingbo big data for providing the dataset for this project. We also thank Meng Xia for her great support. This work is partially supported by the HKUST-SJTU Joint Research Collaboration Fund under Grant No. SJTU20EG02.

References

Ahmad, A. R.; and Murad, H. R. 2020. The impact of social media on panic during the COVID-19 pandemic in Iraqi Kurdistan: online questionnaire study. *JMIR*, 22(5): e19556.

Allgaier, J.; and Svalastog, A. L. 2015. The communication aspects of the Ebola virus disease outbreak in Western Africa–do we need to counter one, two, or many epidemics? *Croatian medical journal*, 56(5): 496.

Andrews, C.; Fichet, E.; Ding, Y.; Spiro, E. S.; and Starbird, K. 2016. Keeping up with the tweet-dashians: The impact of'official'accounts on online rumoring. In *CSCW*, 452–465.

Arif, A.; Robinson, J. J.; Stanek, S. A.; Fichet, E. S.; Townsend, P.; Worku, Z.; and Starbird, K. 2017. A closer look at the self-correcting crowd: Examining corrections in online rumors. In *CSCW*, 155–168.

Arif, A.; Shanahan, K.; Chou, F.-J.; Dosouto, Y.; Starbird, K.; and Spiro, E. S. 2016. How information snowballs: Exploring the role of exposure in online rumor propagation. In *CSCW*, 466–477.

Baumeister, R. F.; Bratslavsky, E.; Finkenauer, C.; and Vohs, K. D. 2001. Bad is stronger than good. *Review of general psychology*, 5(4): 323–370.

Blei, D. M.; Ng, A. Y.; and Jordan, M. I. 2003. Latent dirichlet allocation. *JMIR*, 3: 993–1022.

Brennen, J. S.; Simon, F.; Howard, P. N.; and Nielsen, R. K. 2020. Types, sources, and claims of Covid-19 misinformation. *Reuters Institute*, 7: 3–1.

Bruder, M.; Haffke, P.; Neave, N.; Nouripanah, N.; and Imhoff, R. 2013. Measuring individual differences in generic beliefs in conspiracy theories across cultures: Conspiracy Mentality Questionnaire. *Front. Psychol.*, 4: 225.

Bruno, N. 2011. Tweet first, verify later? How real-time information is changing the coverage of worldwide crisis events. *Reuters Institute for the Study of Journalism*, 2010–2011.

Cacioppo, J. T.; and Berntson, G. G. 1994. Relationship between attitudes and evaluative space: A critical review, with emphasis on the separability of positive and negative substrates. *Psychological bulletin*, 115(3): 401.

Chen, K.; Chen, A.; Zhang, J.; Meng, J.; and Shen, C. 2020. Conspiracy and debunking narratives about COVID-19 origins on Chinese social media: How it started and who is to blame. *Harvard Kennedy School Misinformation Review*.

Chen, L.; Wang, X.; and Peng, T.-Q. 2018. Nature and diffusion of gynecologic cancer–related misinformation on social media: Analysis of tweets. *JMIR*, 20(10): e11515.

Chua, A. Y.; and Banerjee, S. 2017. To share or not to share: the role of epistemic belief in online health rumors. *International journal of medical informatics*, 108: 36–41.

Chua, A. Y.; and Banerjee, S. 2018. Intentions to trust and share online health rumors: An experiment with medical professionals. *Comput Hum Behav*, 87: 1–9.

Cullen, R. 2006. *Health information on the internet: A study of providers, quality, and users*. Greenwood Publishing Group.

De Choudhury, M.; Kumar, M.; and Weber, I. 2017. Computational approaches toward integrating quantified self sensing and social media. In *CSCW*, 1334–1349.

DiFonzo, N. 2008. *The watercooler effect: A psychologist explores the extraordinary power of rumors*. Penguin.

DiFonzo, N.; and Bordia, P. 2007. *Rumor psychology: Social and organizational approaches.* American Psychological Association.

DiFonzo, N.; Robinson, N. M.; Suls, J. M.; and Rini, C. 2012. Rumors about cancer: Content, sources, coping, transmission, and belief. *J Health Commun*, 17(9): 1099–1115.

Dredze, M.; Broniatowski, D. A.; and Hilyard, K. M. 2016. Zika vaccine misconceptions: A social media analysis. *Vaccine*, 34(30): 3441.

Dutta, S.; Ma, J.; and De Choudhury, M. 2018. Measuring the impact of anxiety on online social interactions. In *ICWSM*.

Freeman, D.; Waite, F.; Rosebrock, L.; Petit, A.; Causier, C.; East, A.; Jenner, L.; Teale, A.-L.; Carr, L.; et al. 2020. Coronavirus conspiracy beliefs, mistrust, and compliance with government guidelines in England. *Psychol. Med.*, 1–30.

Ghenai, A. 2017. Health misinformation in search and social media. In *ICDH*, 235–236.

Ghenai, A.; and Mejova, Y. 2017. Catching Zika Fever: Application of Crowdsourcing and Machine Learning for Tracking Health Misinformation on Twitter. In *2017 IEEE International Conference on Healthcare Informatics*, 518–518. IEEE.

Granger, C. W. 1969. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, 424–438.

Gui, X.; Kou, Y.; Pine, K.; Ladaw, E.; Kim, H.; Suzuki-Gill, E.; and Chen, Y. 2018. Multidimensional risk communication: public discourse on risks during an emerging epidemic. In *CHI*, 1–14.

Gui, X.; Kou, Y.; Pine, K. H.; and Chen, Y. 2017a. Managing uncertainty: using social media for risk assessment during a public health crisis. In *CHI*, 4520–4533.

Gui, X.; Wang, Y.; Kou, Y.; Reynolds, T. L.; Chen, Y.; Mei, Q.; and Zheng, K. 2017b. Understanding the patterns of health information dissemination on social media during the Zika outbreak. In *AMIA Annu. Symp. Proc.*, volume 2017, 820. American Medical Informatics Association.

Islam, M. S.; Sarkar, T.; Khan, S. H.; Mostofa Kamal, A.-H.; Hasan, S. M. M.; Kabir, A.; Yeasmin, D.; Islam, M. A.; Amin Chowdhury, K. I.; Anwar, K. S.; et al. 2020. COVID-19–Related Infodemic and Its Impact on Public Health: A Global Social Media Analysis. *The American journal of tropical medicine and hygiene*, 103(4): 1621.

Kou, Y.; Gui, X.; Chen, Y.; and Pine, K. 2017. Conspiracy talk on social media: collective sensemaking during a public health crisis. *CSCW*, 1(CSCW): 1–21.

Lewandowsky, S.; Ecker, U. K.; Seifert, C. M.; Schwarz, N.; and Cook, J. 2012. Misinformation and its correction: Continued influence and successful debiasing. *Psychological science in the public interest*, 13(3): 106–131.

Liao, Q.; and Shi, L. 2013. She gets a sports car from our donation: rumor transmission in a chinese microblogging community. In *CSCW*, 587–598.

Lu, Z.; Jiang, Y.; Lu, C.; Naaman, M.; and Wigdor, D. 2020. The Government's Dividend: Complex Perceptions of Social Media Misinformation in China. In *CHI*, 1–12.

Lütkepohl, H. 2005. *New introduction to multiple time series analysis*. Springer Science & Business Media.

McPherson, M.; Smith-Lovin, L.; and Cook, J. M. 2001. Birds of a feather: Homophily in social networks. *Annual review of sociology*, 27(1): 415–444.

Miller, G. 2020. Researchers are tracking another pandemic, too—of coronavirus misinformation. In *American Association for the Advancement of Science*.

Nyhan, B.; Reifler, J.; Richey, S.; and Freed, G. L. 2014. Effective messages in vaccine promotion: a randomized trial. *Pediatrics*, 133(4): e835–e842.

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

Ortutay, B.; and Klepper, D. 2020. Virus outbreak means (mis) information overload: How to cope. *ABC News*.

Ozturk, P.; Li, H.; and Sakamoto, Y. 2015. Combating rumor spread on social media: The effectiveness of refutation and warning. In *HICSS*, 2406–2414. IEEE.

Pal, A.; Chua, A. Y.; and Goh, D. H.-L. 2019. Debunking rumors on social media: The use of denials. *Comput Hum Behav*, 96: 110–122.

Petty, R. E.; and Cacioppo, J. T. 2012. *Communication and persuasion: Central and peripheral routes to attitude change*. Springer Science & Business Media.

Rajdev, M.; and Lee, K. 2015. Fake and spam messages: Detecting misinformation during natural disasters on social media. In *WI-IAT*, volume 1, 17–20. IEEE.

Röder, M.; Both, A.; and Hinneburg, A. 2015. Exploring the space of topic coherence measures. In *WSDM*, 399–408.

SCMP. 2020. China's mask shortage hits global supply as outbreak rages on. https://www.scmp.com/economy/china-economy/article/3050717/coronavirus-chinas-surgical-mask-shortage-ripples-through. Accessed: 2020-09-01.

Sell, T. K.; Hosangadi, D.; and Trotochaud, M. 2020. Misinformation and the US Ebola communication crisis: analyzing the veracity and content of social media messages related to a fear-inducing infectious disease outbreak. *BMC Public Health*, 20: 1–10.

Shin, J.; Jian, L.; Driscoll, K.; and Bar, F. 2017. Political rumoring on Twitter during the 2012 US presidential election: Rumor diffusion and correction. *new media & society*, 19(8): 1214–1235.

Smailhodzic, E.; Hooijsma, W.; Boonstra, A.; and Langley, D. J. 2016. Social media use in healthcare: a systematic review of effects on patients and on their relationship with healthcare professionals. *Health Serv Res*, 16(1): 442.

Starbird, K.; Dailey, D.; Mohamed, O.; Lee, G.; and Spiro, E. S. 2018. Engage early, correct more: How journalists participate in false rumors online during crisis events. In *CHI*, 1–12.

Starbird, K.; Maddock, J.; Orand, M.; Achterman, P.; and Mason, R. M. 2014. Rumors, false flags, and digital vigilantes: Misinformation on twitter after the 2013 boston marathon bombing. *IConference 2014 Proceedings*.

Tanaka, Y.; Sakamoto, Y.; and Matsuka, T. 2013. Toward a social-technological system that inactivates false rumors through the critical thinking of crowds. In *HICSS*, 649–658.

Tasnim, S.; Hossain, M. M.; and Mazumder, H. 2020. Impact of rumors and misinformation on COVID-19 in social media. *J Prev Med Public Health*, 53(3): 171–174.

Van Prooijen, J.-W.; and Jostmann, N. B. 2013. Belief in conspiracy theories: The influence of uncertainty and perceived morality. *Eur J Soc Psychol*, 43(1): 109–115.

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

Walter, N.; Brooks, J. J.; Saucier, C. J.; and Suresh, S. 2020. Evaluating the impact of attempts to correct health misinformation on social media: A meta-analysis. *Health Commun*, 1–9.

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

Weibo, S. 2020. SEC Filing of Weibo Corporation. http://ir.weibo.com/node/7726/html. Accessed: 2020-09-01.

Yang, D.; Halfaker, A.; Kraut, R.; and Hovy, E. 2016. Who did what: Editor role identification in Wikipedia. In *ICWSM*, volume 10, 446–455.

Zarocostas, J. 2020. How to fight an infodemic. *The Lancet*, 395(10225): 676.

Zhang, X.; Liu, S.; Deng, Z.; and Chen, X. 2017. Knowledge sharing motivations in online health communities: A comparative study of health professionals and normal users. *Comput Hum Behav*, 75: 797–810.