What drives people to repost social media messages during the COVID-19 pandemic? Evidence from the Weibo news microblog

Jiayin Pei1 | Zhi Lu2 | Xiaoming Yang3

1School of Business, Jiangnan University, Wuxi, China
2Peter B. Gustavson School of Business, University of Victoria, Victoria, British Columbia, Canada
3College of Business Administration, University of Nebraska at Omaha, Omaha, Nebraska, USA

Correspondence Xiaoming Yang, College of Business Administration, University of Nebraska at Omaha, Omaha, NE 68182, USA. Email: xyang2@unomaha.edu

Funding information National Social Science Foundation of China, Grant/Award Number: 20FTQ018; JiangSu Province Social Science Foundation, Grant/Award Number: 20TQB001; The Fundamental Research Funds for the Central Universities, Grant/Award Number: 2019JDZD06

Abstract COVID-19 poses an unprecedented challenge to human society. To cope with the pandemic, people seek information from various communication channels. Microblog websites are highly influential information channels for the public to get timely information during the pandemic. Building on the heuristic-systematic processing model, this study identifies the multiple characteristics (content, author, and social features) that may play a role in triggering long cascades of reposts of COVID-19-related news microblogs. With a large-scale news microblog database collected from Weibo and an innovative information gain method, we find that heuristic thinking plays a dominant role in COVID-19 pandemic-related news microblog reposting decisions and further discloses the specific influencing factors of such behavior.

1 | INTRODUCTION

COVID-19 has been the most salient and influential public health crisis worldwide since the beginning of 2020. Due to the urgency, uncertainty, and the continually changing nature of the pandemic, numerous news, facts, and rumors spread via all kinds of channels such as TV, radio, newspaper, and social media. Microblogging sites are emerging as a fast communication channel for gathering and spreading pandemic-related information (Hasan et al., 2019; Ucar et al., 2018; Zheng et al., 2019). In the throes of a pandemic, social media, including microblogging, could disseminate up-to-date information about epidemic prevention and influence public attitudes and behaviors (Wang et al., 2020; Wiederhold, 2020). It could also become the source of viral misinformation and even exacerbate the panic if not managed well (Baum et al., 2020; Larson, 2018).
Therefore, analyzing this rich microblog content can yield valuable new knowledge in understanding the pattern and influencing factors of information dissemination through social media that enables social media platforms and policymakers to manage their information dissemination and cope with the pandemic in a more effective manner (Akcura et al., 2018).

Extant research on the influence factors of microblog dissemination mainly focuses on studying microblog content characteristics and user characteristics on reposting behavior. In the microblog content characteristics domain, prior research has produced inconsistent findings such that some studies find longer messages are more persuasive than shorter messages. In contrast, other studies report no such effect (Guan et al., 2014). Meanwhile, some studies only analyze word features through text mining, which has proven to be inefficient and leads to findings that are hard to interpret (Carrera & Jung, 2018). In the domain of examining the factors that affect microblog users’ forwarding behavior, it is suggested that people are more likely to pay attention to those professional, active, or experienced Weibo users and only read the Weibo posted by those users (Gao et al., 2015). Recent studies unveil that both microblog user characteristics and content attributes jointly affect the dissemination of microblog information (Lu et al., 2014). However, there is plenty of room to improve in this stream of research. For example, some research uses the interview method to create factors, yet these factors are not applicable for automatically detecting public perceptions that are essential for user behavior prediction tasks (Ormand & Warkentin, 2015). Further, some research use factors collected directly from API (Application Programming Interface), which is too simple to represent the diverse features of a post (Wu et al., 2019). In a similar vein, previous work investigates the influencing factors of the forwarding behavior of users that are related to the blog author but has not considered other integral factors, including user experience and authority.

In addition, extant studies generally assume that microblog users independently evaluate the news without considering other users’ social impact on a single user (Gursoy, 2019). Nonetheless, this assumption overlooked a deeply ingrained feature of social media, that is, its unique “following” relationships among users and the social connections it fosters. People can spread their wisdom and experience knowledge in the social network space through “reposting” (retweeting) behavior. In light of this, people are unlikely to make decisions individually; instead, they are very likely to use the collective wisdom of the public as inspiration to evaluate microblog news (Cocosila & Igonor, 2015). A few scholars have gradually begun to study the role of social impact in microblog news reposting (Zhang et al., 2014). To date, most studies only characterize social impact through the structure of social networks. Few predictive models integrate the behavior characteristics of other users and microblog authors with other content attributes of microblogs to predict the trend of these microblogs spreading (Wu & Shen, 2015).

To sum up, this research aims to address these gaps by providing a more in-depth and thorough understanding of the influencing factors of microblog news dissemination during a health crisis such as the COVID-19 pandemic. We find that social influence characteristics are the most dominant factor influencing the repost and spread of Weibo microblog information, followed by author characteristics and content characteristics. We further shed light on this finding and theorize that this effect is due to microblog users’ heuristic information processing within the vast amount of social media communication context. In doing so, our research makes three main contributions to the extant literature. First, we draw from the dual processing theory model, persuasion theory, and social influence theory to delineate the whole spectrum of microblog characteristics that potentially influence microblog reposting behavior. Second, we scrape a large-scale COVID-19-related microblog data set from Weibo, the largest and most influential social media platform in China, and innovatively employ the
information gain method to compare the importance of multiple microblog characteristics on reposting behavior empirically, including the social influence of microblog, the characteristics of microblog content, and the characteristics of microblog authors. Third, our research leverages the Weibo platform to collect first-hand social media communication big data among its substantial number of users and consequently track how the information spreads as the COVID-19 crisis unfolds, which we believe exemplifies the microblog communication behavior of news media and illuminates the news dissemination pattern in the public sector. Taken together, this current research discloses intriguing insights into social media users’ reposting (retweeting) decision process during the COVID-19 pandemic that deepens the understanding of the users’ reposting behaviors of microblog messages and contributes to both the body of literature on social media communication and the practice of communication management in the era of social media.

2 | THEORETICAL BACKGROUND

Microblog users participate in discussing information related to COVID-19 by posting their news or reposting others’ news microblogs. Reposting accounts for a majority of the news messages people share on social media platforms. Therefore, reposting behavior largely determines the spread of news communication on these platforms. Reposting behavior is a result of various influencing factors such as specific attributes contained in the microblog, the characteristics of the microblog author, and how other users respond to the blog, among others. To better understand the information processing process of microblog users when making reposting decisions, we draw from the dual processing model, persuasion theory, and social influence theory and use this framework to gain a comprehensive understanding of the factors affecting the spread of microblog news during the pandemic.

2.1 | Dual processing theory

“Dual processing theory” is the basic theory explaining the cognitive processing process of human beings (also named as Heuristic-Systematic Model, Chaiken, 1980). It distinguishes two information processing modes: systematic information processing and heuristic information processing. The systematic information processing process denotes people expend more mental effort during the process and have the will and motivation to collect information (Khadjah Nassirtoussi et al., 2014). The systematic processing process relies on rules and logic, and therefore, it is comprehensive and slow. In contrast, the heuristic processing process refers to the fact that human beings cannot perform comprehensive and systematic analysis following the rules in a complex environment. Rather, they can only resort to experience and intuition to process information. During this process, people have to ignore some clues to reduce their cognitive burden and therefore use less time (Tversky & Kahneman, 1974, 1981). As Figure 1 shows, the dual processing model illustrates that quick and simple intuitive (heuristic) decision making and deliberate and rational (systematic) decision making are not made separately; on the contrary, the two processing methods have an effect on the human decision-making process simultaneously and work in parallel. Generally speaking, the advantage of the systematic processing system is that it conforms to the rules and can guarantee the accuracy of the decision-making results; the advantage of the heuristic processing system is that it needs low cognitive effort and has a rapid response.
2.2 | **Persuasion theory**

Although previous studies have concluded that the content (Neuman, 1997) attributes of microblogs will affect the scale of reposting, yet one thing that stands out from these studies is each study has different methods for measuring the concept of content attributes of microblog. Persuasion theory suggests that two important factors affect a person’s decision-making process: the persuasion strength contained in the information and the persuasion attitude (Angst & Agarwal, 2009). We reason that persuasion theory can be used to conceptualize the influencing factors of microblog content attributes and further measure them.

Strong persuasion can guide the thinking of others and further effectively persuade and influence the decision making and behavior of others (O’Keefe, 2015). In the social media environment, the persuasive power contained in microblog content can influence the information dissemination decision of microblog users. The second important content factor that affects users’ decision making is persuasive attitude (positive or negative), that is, emotional factor (Fan et al., 2019; O’Keefe, 2015). Emotions play a significant role in decision-making and information processing, for example, which candidate to support, which product to buy, which piece of information to disseminate, and so forth.

2.3 | **Social influence theory**

Social influence refers to the influence of people’s emotions, opinions, and behaviors on others (Latane, 1996). Social influence theory suggests that the more social participants, the greater the social influence. Social psychology and cognitive science have testified the importance of social influence and put forward different theories (O’Keefe, 2015). For example, the Consensus heuristic indicates that if many people think something is good, others will also tend to hold the same point of view. Similarly, the Bandwagon effect states that to gain recognition or avoid disapproval, people tend to take actions consistent with the majority of people (Simon, 1954). In addition, the theory of reasoned action reveals that if a person believes that most people approve of specific behavior, their behavior will be affected accordingly (Fishbein & Ajzen, 1977).

These theories underpin the influence of many social participants’ behavior on the behavior of other users in the traditional environment. Today’s social media platforms connect people closely and make it more difficult for people to make decisions independently. As a consequence,
the actions taken by most people affect the actions of others. Recent social media research on user behavior and information dissemination shows that social media users do not adopt rational analysis strategies of carefully reading every piece of information during information processing. Instead, they are more inclined to rely on the behavior of other people as their decision cue. Others’ behavior could help them make decisions to reduce the potential negative effects of making decisions different from others. In light of social influence theories, we propose that microblog users will use the social influence characteristics of Weibo microblog as heuristic cues when they make reposting decisions. Therefore, the social influence aspects of microblogs are also an influencing factor of information dissemination.

3 | INFLUENCING FACTORS OF MICROBLOG REPOSTING BEHAVIOR BASED ON DUAL PROCESSING THEORY MODEL

3.1 | Research conceptual model

The dual processing model indicates that the thoughtful and rational systematic processing and the fast and simple heuristic processing work in concert with each other and both affect human decision-making processes. Previous studies have attested to the impact of rational systematic information processing on microblog reposting behavior, such as analyzing the impact of information display form or structural characteristics on reposting behavior (the length of microblog content, the appearance of titles, topics, etc.). At the same time, the influence of heuristic information processing means that people’s decisions will be affected by an array of heuristic clues. For example, both the characteristics of microblog authors (expertise, popularity) and other users’ repostings can be used as heuristic cues and lead microblog users to decide whether to repost the microblog without carefully checking its content.

Therefore, building on the dual processing model, this research categorizes the influencing factors of microblog reposting behaviors into three aspects: microblog content characteristics, microblog author characteristics, and microblog social influence characteristics. The research model is shown in Figure 2.

3.2 | Microblog content characteristics that affect reposting

In the systematic processing mode, microblog users will carefully read and evaluate the microblog content as a basis for their reposting decisions. According to the aforementioned Persuasion Theory, this research conceptualizes the content characteristics of microblogs into the strength and persuasive attitude of microblog posts.

3.2.1 | Microblogging features that reflect persuasion strength

The strength of persuasion is usually expressed by the degree of relevance, timeliness, accuracy, and comprehensiveness (Cheung & Thadani, 2012). Given all the microblog news released by the most influential official media are accurate, this study only proposes the factors affecting the spread of microblog news based on the remaining three aspects and specifies their measurements.
1. Factors reflecting the degree of relevance of microblogs: The degree of relevance between microblog content and emergencies was measured by the frequency of emergence of incident-related words (Pei et al., 2017) in microblog news.

2. Factors reflecting the timeliness of microblogs: Timeliness refers to the priority of microblog news information at the release time, which will affect the persuasiveness of microblogs. Timeliness is measured in two ways in our research. First, compared with other microblog news, the earliness of the release time of microblog news. Second, the release time period of the microblog. For example, microblog news posted at peak times is easier to be seen by other users and is thus more likely to be reposted. In this research, we divided the release time into three periods: 1–8 o’clock, 9–18 o’clock, and 19–24 o’clock.

3. Factors reflecting the comprehensiveness of microblogs: Comprehensiveness refers to the richness of related information described by microblog news. In addition to the richness of microblog news text, title, and topic, it is also affected by text features such as pictures or links to other websites. These nontext features will change the display form of microblog news content and affect persuasiveness as well.

3.2.2 Microblogging features that reflect persuasion attitude

Prior research suggests grouping the emotional characteristics of persuasion attitudes into three categories: lexical characteristics, syntactic characteristics, and emotional symbol characteristics (Anders et al., 2020; Frank et al., 2013; Liu et al., 2016; Zhao et al., 2016). In a similar line, this research proposes the influencing factors of persuasion attitude as shown in Table 1.
1. Vocabulary features: Vocabulary features summarize the features of microblogs from the perspective of characters and words. The positive and negative sentiment vocabulary used in this research is the sentiment vocabulary released by the NTUSD project of National Taiwan University.

2. Syntactic features: Syntactic features include sentences, punctuation features, etc. The characteristics of punctuation marks are also effective in sentiment classification. Some punctuation marks, such as exclamation marks, ellipsis, and question marks, can show the strength of a user’s positive or negative attitude toward a particular point of view.

3. Emotion symbol features: Emotion symbols are also a very important feature to distinguish between positive and negative emotions, such as Cry, Sad, LOL, Puzzled, and Think hard, etc.

### 3.3 Characteristics of microblog authors that affect reposting

Although an in-depth analysis of news information can bring higher decision-making accuracy, in the face of information overload, microblog users may not read each microblog carefully and

| TABLE 1 | Influencing factors of microblog news content characteristics |
|----------|---------------------------------------------------------------|
| **Item** | **Description**                                               |
|Persuasion strength | |
|Relevance characteristics | |
|1–3 | FS/Med/EP WordCount | The number of keywords containing “food safety,” “medical,” and “environmental pollution” in Weibo |
|4 | Hour of Day | Weibo release time |
|5 | Hour Range | Weibo release time period (1–8) (9–20) (21–24) |
|Comprehensiveness characteristics | |
|6–9 | Title/Topic/URL/Pic | Does Weibo contain titles, topics, external links, pictures |
|10 | At | Does Weibo mention others (“@someone”) |
|11–12 | Text/Title Length | Weibo, Weibo title length |
|Persuasion attitude | |
|Lexical characteristics | |
|13–15 | Wordonce/twice/three | Number of Weibo keywords that appeared 1, 2, or 3 times in Weibo |
|16–17 | Num/Eng Count | Weibo contains numbers and numbers in English |
|18–19 | UniPOS/NEG WordCount | Number of positive and negative emotion words appearing in Weibo |
|20 | Wordfre | Number of Weibo keywords |
|Emotional symbol features | |
|21–23 | Posi/Neu/NegFace Count | Number of positive, neutral, and negative emoticons on Weibo |
|Syntactic features | |
|24–25 | Sentence/Punction Count | Weibo contains sentences and punctuation marks |
|26 | SentiMarkCount | Number of punctuation marks in Weibo that reflect emotional strength |
make reposting decisions based on the content of the microblog. Therefore, people do not simply perform rational analysis and systematic processing when making decisions; as a matter of fact, they must also perform heuristic information processing. In microblog reposting decision making, users may judge which microblog is more worthy of reposting based on the following two heuristic cues.

1. The popularity heuristic: Gigerenzer (1996) first proposed the concept of the popularity heuristic. The popularity heuristic means that between the two options if the decision maker is familiar with one option but not the other, the decision maker will think that the one they are familiar with has a higher value. Familiarity is the criterion for choosing the “familiarity heuristic” strategy. Microblog users are more likely to be familiar with those active and experienced users.

2. Authoritative heuristics: Due to the attractiveness of news information sources, microblog users may be willing to repost microblogs published by influential microblog users. They assume that those authoritative users’ or experts’ information are worth reposting.

In summary, microblog users are more willing to repost microblog posts by active users, experienced users, or authoritative users. Consequently, this proposes the characteristics of microblog authors that affect the reposting of microblogs from these three aspects (see Table 2) as follows:

1. Factors of a microblog that reflect the author’s activity: Activity on a microblog often means the maturity of the user’s social relationship, which will promote the reposting of a microblog. A large number of microblog posts since an account’s registration indicates that the user has continually active performance on the microblogging platform. With the guarantee of the number of microblog posts, the possibility of the microblog being reposted will also increase. Personal experience data such as the author’s gender and location can also reflect the difference in the author’s activity.

2. Factors of a microblog that reflect the author’s seniority: Users who have used microblogs for a long time tend to understand the rules of the microblogging platform better, build a good interaction record with friends, and have accumulated high proficiency to use the microblog as a social tool.

3. Factors of a microblog that reflect the author’s authority level: In line with authoritative heuristics, microblog users are more willing to forward microblog posts published by authoritative microblog users (Yin et al., 2020). The influencing factors that reflect the author’s degree of authority include:
   a. Author’s authentication information: To ensure the authenticity of microblog users, users who have been officially authenticated by the Weibo microblogging platform will display their logos of authentication. Meanwhile, certified users who want to increase their degree of popularity on the microblogging platform will devote themselves to publishing rich microblogs, which also boosts their levels of authentication.
   b. Author’s personal authentication details: The personal details of the authenticated user mainly include information about the user’s occupation, professional contribution, etc. The more detailed the personal authentication details, the more people tend to think that the source of their microblog information is more authoritative and reliable.
   c. Author’s personal profile: A microblog author’s personal page usually displays his/her personal profile, mainly including occupation, interests, work focus, and other information.
These contents indicate the author's personal information and enhance people's understanding of the author's authority.

d. Author’s influence: Many social media platforms publish the influence index of each news blogger (author). The ranking of news bloggers in the list will affect other users’ perception of the authority of a news author.

e. The number of a microblog author’s fans: Major social media platforms all have a similar fan mechanism. For example, microblog users can independently choose from whom to follow and receive news and information. The number of fans or followers acts as another indicator of the microblog author’s degree of authority.

### 3.4 Social impact characteristics of a microblog that affect reposting

From a sociological point of view, the public has a herd mentality and they are more inclined to be influenced by their peers and participate in a specific behavior (Simon, 1955;
Yin et al., 2020). The same is true in the microblogging platform. The number of times each microblog has been reposted is clearly shown on the platform. The reposted number implies the degree of attention that a particular microblog has received. Suppose microblog users find that a certain number of users have reposted a certain microblog. In that case, their confidence in reposting this blog will increase, which further stimulates more users to repost this microblog.

Therefore, this research also takes into account the social influence characteristic of the number of reposts in the initial microblog release stage on the user’s reposting decision. More specifically, we divide the first 2 hr of microblog release into six intervals and count the number of reposts in the first 20, 40, 60, 80, 100, and 120 min of microblog posting as the social impact characteristics of the microblog (Table 3). Doing so could also help us find the optimal timing of social media communication’s amplifying effect.

4 | METHOD AND DATA

4.1 | Data

The research data were collected from Weibo, the leading microblogging platform in China that has over 511 million monthly active users as of September 2020. Our data set consists of the microblogs continuously published by 12 mainstream news outlets and 10 government microblog accounts covering public health issues. These 22 mainstream news outlets and government microblog accounts were selected according to the “purposive” sampling method (Chris et al., 2002), which states two criteria for media content analysis: focusing on the key media and selecting the most relevant media. We followed these two criteria to identify online news media and government accounts in two steps. The first step is to identify a diversity of influential news sources. As a result, we selected 24 top 20% news outlets from different media types including newspaper, magazine, radio, TV, and website. We also chose 22 top 20% government

| TABLE 3 | Descriptive statistics of microblog users |
|----------|----------|----------|----------|----------|
|           | Min      | Max      | Mean     | SD       |
| author_statuses_count | 12,949    | 89,516   | 34,713   | 242,03l  |
| author_friends_count | 25        | 1,975    | 410.182  | 398.385  |
| author_favourites_count | 0        | 1,762    | 249.909  | 517.540  |
| DaysActive | 1,766     | 4,061    | 3,571.409| 553.127  |
| author_bifollowers_count | 7        | 1,298    | 221.636  | 280.571  |
| Last year  | 625       | 23,104   | 8,239.545| 8,599.218|
| orepcount  | 37,221    | 37,800,966| 6,680,321| 10,988,400|
| Top1SumRep | 737       | 287,325  | 62,467.727| 80,638.080|
| Descriptionlen | 9      | 85       | 44.182   | 25.558 |
| Verifyreasonlen | 4      | 19       | 10.409   | 4.404   |
| InfluenceScore | 82.37   | 97.524   | 91.072   | 4.754   |
| author_followers_count | 1,535    | 33,700,000| 6,178,858.409| 8,806,591.959|
microblog accounts from different levels of government, including central government and province, city, and county levels of government in different regions of China. The second step is to delete the irrelevant microblog accounts. We manually deleted those news media and accounts focusing primarily on economics, entertainment, tourism, education, sports, and traffic which left us 22 microblog accounts. Among all the public health issues microblog posts, we randomly selected 1,000 microblog posts that contain “New Coronavirus,” “New Coronary Pneumonia,” “COVID-19,” or “New Coronary Epidemic.”

To analyze the reposting and dissemination trend better to understand the public concern about the COVID-19 pandemic, we traced all 735,441 reposts of the original 1,000 posts. Figure 3 shows a distribution of the number of reposts published by the 22 news outlets/government microblog accounts in the previous year such that those microblogs published by People’s Daily and CCTV News have more repostings than other news outlets. Likewise, Chinese Police Online, China Fire Protection, and China Peace have more reposts than other government microblog accounts.

According to previous research, we set the threshold of reposting to 766 (Pei et al., 2017) and divided the 1,000 microblog posts into two groups. The 321 posts that were reposted 766 times or more were defined as widely disseminated posts and categorized as Group 1, whereas the remaining 679 posts were defined as normal posts and categorized as Group 2. Then, we compared the two groups to examine the influencing factors of the microblogging feature with an innovative method of information gain. Doing so helps us to identify those features that have an important influence on COVID-19 pandemic-related reposting, which is of great interest to social media communication and public opinion practitioners.

**FIGURE 3** Distribution of the number of reports published by the 22 microblog accounts
4.2 Selection of influencing factors based on information gain of microblogs

Prior research has attested that not all of the influencing factors we proposed from the perspective of the dual-processing model help predict microblog users’ reposting behavior. For example, an increase in the length of microblog content will help people better understand the information related to emergencies and increase microblog dissemination. However, if the information’s content is too long, it will reduce the popularity of reposting. This challenge makes the logistic regression and structural equation modeling methods that are widely used in the prior research of influencing factors not suitable for this current research. In light of this, we decide to employ the information gain method to measure how much information a known feature (characteristic) contributes to microblog users’ reposting behavior (Abbasi et al., 2008). More specifically, we measure the information entropy of the classification result by judging the presence and absence of a specific feature. The measurement method is described in detail below.

Let \( \mathcal{L} \) be a learning sample of microblogs, \( \mathcal{L} \supseteq \{(X_1, c_1), \ldots, (X_m, c_j)\} \). We denote \( m \) by the number of microblogs in \( \mathcal{L} \) \( \forall i \in \{1, \ldots, m\} \), \( X_i \) is a measurement vector of microblog \( i \), \( X_i \in \mathcal{M} \) \( \mathcal{M} \) being the measurement space. \( \forall i \in \{1, \ldots, J\} \), \( c_i \) represents category \( i \), and \( c_p \in Z \) where \( Z = \{c_1, c_2, \ldots, c_k\} \) is the set of categories. \( ||c|| \) denotes the number of microblogs of \( \mathcal{L} \) that are in category \( c_i \). The prior probability that a microblog belongs to a given category \( c_i \) is given by \( P(c_i) = \frac{||c||}{m} \). The components of the vector \( X_i \) can be viewed as factors of a microblog post that are \( x_1, x_2, \ldots, x_p \), and a test on the influencing feature is based on one of these features. Given a test \( T \) (based on a single feature \( x_i \)), with \( n \) possible feature values, we denote by \( t_i \) the set of the microblogs in \( \mathcal{L} \) that are in category \( c_i \). The probability that test \( T \) has the feature value \( i \) is estimated by \( P(t_i) = \frac{||c_i, t_i||}{n} \). The number of microblogs of \( \mathcal{L} \) that are in category \( c_i \) and have the feature value \( j \) for test \( T \). The probability that an object is in \( c_i \) and has the feature \( j \) is given by \( P(c_i, t_j) = \frac{||c_i, t_j||}{n} \). The conditional probability, \( p(c_i|t_j) \), that an object is in the category \( c_i \) under the condition that the test \( T \) has the feature value \( j \), is estimated by \( \frac{P(c_i, t_j)}{P(t_j)} \). Obviously, we have:

\[
\sum_{i=1}^{k} P(c_i|t_j) = 1 \forall j \in \{1, \ldots, n\} \cdot P(c_i) \cdot P(c_i|t_j) \cdot P(t_j) \in [0, 1] \text{ and } P(c_i|t_j) = \frac{P(c_i, t_j)}{P(t_j)} \forall j \in \{1, \ldots, n\} \text{ and } \forall i \in \{1, \ldots, k\}.
\]

The Information Gain has its origin in information theory (Raileanu & Stoffel, 2004). It is based on the notion of entropy that characterizes the uncertainty of an arbitrary set of examples. If we randomly select a microblog from a learning sample of microblogs \( \mathcal{L} \) and we announce that it belongs to category \( c_i \), then the probability of this message is equal to \( P(c_i) = \frac{||c||}{m} \), and the amount of information it conveys is \( - \log_2 P(c_i) \). The logarithm is in base 2 because the entropy is a measure of the expected encoding length measured in bits. The expected information provided by a message with respect to the set of categories \( Z \) can be expressed as formula (1):

\[
H(Z) = - \sum_{i=1}^{k} P(c_i) \log_2 P(c_i)
\]

The quantity \( H(Z) \) measures the average amount of information needed to identify the category of a microblog in \( \mathcal{L} \). This quantity is also known as the entropy of the set \( \mathcal{L} \) relative to the \( k \)-wise classification. In the context of using this method to analyze the impact of different microblogging features on reposting, we divide \( Z \) into \( c_1 \) (public hotspot microblogs) and \( c_2 \)
(ordinary microblogs), where \( x_1, x_2, ..., x_n \) are microblog characteristics that affect the reposting of a microblog. We restrict the information entropy to the situation in which we have only two possible categories \( k = 2 \). Therefore, we have:

\[
H(Z) = - \sum_{i=1}^{2} P(c_i) \log_2 P(c_i)
\]  

We will consider a similar measurement on a test \( T \) with the presence of a single feature \( x_i \). Assuming the feature \( x_i \) has \( n \) values, the expected information requirement, also called the conditional information entropy, is the weighted sum over each feature value:

\[
H(Z|x_i) = \sum_{i=1}^{n} \frac{H(Z|c_i)}{P(x_i)} P(x_i)
\]

Since entropy measures the uncertainty of the set \( Z \) relative to the set of categories \( Z \), that is, the change in the amount of information before and after observing feature \( x_i \). The difference in the amount of information before and after is the uncertainty that feature \( x_i \) reduces for the classification. Therefore, the information gain of feature \( x_i \) is defined in formula (4):

\[
IG(Z,x_i) = H(Z) - H(Z|x_i)
\]

The information gain method measures the importance of a feature through the amount of information that the feature can bring to a classification. The greater the amount of information provided, the lower the classification uncertainty, and the more important the feature.

After calculating the information gain value that identifies the importance of the input feature, all the influencing factors of the microblogging feature are sorted by comparing the size of the gain value. The microblogging feature that has an information gain value greater than a certain threshold is a feature that has an important influence on reposting. The larger the value of information gain, the more important the microblogging feature is to influence the prediction of reposting a microblog. We, therefore, take these important microblogging features as factors that influence reposting.

To analyze the impact of different microblogging features on reposting, this study calculated the information gain value of each Weibo microblogging feature on its subsequent reposting. This helps determine the impact of each Weibo feature on COVID-19 news communication and the pandemic public information spreading trend prediction. The calculation of information gain was conducted using WEKA v. 3.6 (Hall et al., 2009). WEKA is a data mining software that includes a collection of state-of-the-art data preprocessing tools and learning algorithms. The information gain algorithm utilized in this research is also included in this software. This research selects 15 microblogging features that have an information gain >0.05 that have an important impact on the reposting of a COVID-19 pandemic microblog. Table 4 lists all Weibo microblogging features selected in descending order with an information gain >0.05.
5 | RESULTS

The results of this research demonstrate that social influence has an important impact on the spread of COVID-19 pandemic microblog information. Table 4 shows that the information gain (IG > 0.598) of the early forwarding characteristics of a Weibo microblog based on social influence is significantly higher than other influencing factors (IG ≤ 0.432), which suggests that social influence characteristics are the most important factor influencing the repost and spread of Weibo microblog information. The research results support Latané’s (1996) theory of social influence and confirm the importance of social influence in social media communication. It is worth noting that within the first 2 hr of the Weibo microblog release, the information gain of Weibo’s social influence features (features 1–6) gradually increases with time. Therefore, within the first 2 hr of the Weibo microblog release, the longer the time, the better the information communication effect.

We postulate that there are three reasons why social influence plays a crucial role in microblog reposting on Weibo. First, a large number of reposts on Weibo means that the quality of Weibo content is higher. Second, a large number of reposts in the early stage of Weibo’s release means that the Weibo microblog has reached a certain range of dissemination that enables more users to see this news microblog and increases its exposure. Finally, this effect is partly due to the recommendation system that recommends Weibo content. More specifically, the recommendation system is based on the popularity of a particular microblog, so the microblog with a high number of early reposts occupies a prominent position in the major news recommendation systems. Consequently, more users see the top news microblog and are more likely to repost them.

Author characteristics are the second most important factor affecting the spread of COVID-19 pandemic news, including those features that show high information gain. More pertinent to this research, we found the reposted number of the author in the previous year, the number of fans

| Rank | Weibo microblogging features                                      | Information gain |
|------|-------------------------------------------------------------------|------------------|
| 1    | Number of retweets received in the first 120 min of Weibo posting| 0.747            |
| 2    | Number of retweets received in the first 100 min of Weibo posting| 0.729            |
| 3    | Number of retweets received in the first 80 min of Weibo posting | 0.675            |
| 4    | Number of retweets received in the first 60 min of Weibo posting | 0.634            |
| 5    | Number of retweets received in the first 40 min of Weibo posting | 0.612            |
| 6    | Number of retweets received in the first 20 min of Weibo posting | 0.598            |
| 7    | Number of times the author’s original Weibo was reposted in the previous year | 0.432 |
| 8    | Author fans                                                      | 0.412            |
| 9    | The author’s highest number of reposts on a single Weibo in the previous year | 0.411 |
| 10   | Author influence                                                 | 0.377            |
| 11   | Author’s province                                                | 0.302            |
| 12   | Weibo title length                                               | 0.287            |
| 13   | Does Weibo contain topics                                        | 0.219            |
| 14   | Number of Weibo posted by author                                 | 0.167            |
| 15   | Does Weibo contain pictures                                      | 0.106            |
of the author, the highest number of reposts of a single Weibo microblog by the author in the previous year, the author’s influence, the author’s province, the author’s activity, seniority, and authority all contribute to the reposting of an author’s microblog. Our results differ from prior literature (Fan et al., 2019; Frank et al., 2013) in the valence (sentiment) of the news does not affect reposting. We suppose that the reason might be the language style (publishing behavior) of the news media account and the public sector account are different from that of regular users. For instance, the content of news outlets and government microblogs and the tone of these microblogs are usually neutral and do not have too many emotional words and expressions.

In this study, the social influence from the heuristic information processing process and the characteristics of Weibo authors provide relatively high information gains for the spread of COVID-19 news, whereas the content characteristics derived from the systematic processing only provide a small amount of information gain. We conclude that the impact of systematic processing on microblog information dissemination is relatively weak. When emergencies occur, the information overload caused by the vast amount of similar information generated in the social media environment may be the reason why Weibo microblog content loses its impact. We reason that this is because people are required to respond as soon as possible under unprecedented challenging emergencies, which leads to increased cognitive load and weakens their systematic information processing; meanwhile, their heuristic processing is basically unaffected. As a result, heuristic analysis processing plays a significant role.

6 | PRACTICAL IMPLICATIONS

This research provides the following implications based on the analysis of factors affecting the spread of microblog news on Weibo in the context of the COVID-19 pandemic.

First, the role of social influence becomes more important in the era of social media. Our research finds that reposting decisions of microblogging platform users are not made independently but are affected by other users. Although the news media determines the content of news published on the platform, it is the reposting behavior of users that determines the spread of news. This poses an interesting question for news media such that their news impact will rely not only on the professional level of their reporters and editors who craft the content but also more heavily on the normal users and general public to repost and spread their news report. This shift in influence source signals the watershed moment of the communication industry. The long-term implications for the media ecosystem and its industry structure deserve future research endeavor.

Second, policymakers are well-advised to leverage social media channels (e.g., microblogging platforms) especially the influential news media to help the public contain the pandemic and get back to normal. Research has confirmed the influence of author characteristics on the spread of news microblogs. Therefore, public administrators like governors and mayors could communicate their response plans or calls for action through the social media channels of influential news media so that their important measures such as practicing social distancing could reach a larger audience and assuage public concern on the pandemic (Chernozhukov et al., 2020).

Third, communication practitioners will probe deep into the analysis of factors affecting the spread of various types of microblogs. Research has suggested that the content characteristics of different types of microblogs have different effects on their spreading and reposting. People may read certain types of microblogs very carefully to get accurate information, whereas for other microblogs, they may simply browse through and do not make much cognitive effort. For example, people usually do not want to put a lot of effort into reading entertainment news.
microblogs. Future research on social media information communication is encouraged to give more attention to the nuanced differences among different types of microblogs and manage them accordingly.

Fourth, the influencing factors studied in this paper lay the foundation for the research of the news dissemination trend prediction model and data analytics in communication research. Given the relationships among these influencing factors are complex and dynamic, the prediction model’s accuracy and generalizability will be undermined if we add all influencing factors to the prediction model. Analyzing and reducing the influencing factors will be constructive to a useful news dissemination prediction model.

7 | CONCLUSION

Drawing on the dual processing model, persuasion theory, and social influence theory, this research conceptualizes the information processing mechanisms (heuristics and systematic processing) in the social media news communication process and analyzes the reposting decisions of large-scale microblog Weibo users during the COVID-19 pandemic. With a novel research method of information gain, we examine the impact of various factors of microblogs on users’ reposting behavior. Our empirical research results disclose that when an emergency occurs, the vast amount of information in the social media environment makes the systematic processing less salient. As a result, microblog users pay less attention to the importance of blog content features. Instead, the characteristics of authors and social features aligned with the heuristic processing process exert more impact on the users’ reposting behavior.

ACKNOWLEDGMENTS
This research is supported by the postfunded project of the National Social Science Foundation of China (grant no. 20FTQB018), JiangSu Province Social Science Foundation (grant no. 20TQB001), and the Fundamental Research Funds for the Central Universities (grant no. 2019JDZD06).

DATA AVAILABILITY STATEMENT
The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID
Zhi Lu
https://orcid.org/0000-0001-7625-0573
Xiaoming Yang
https://orcid.org/0000-0002-2626-2223

REFERENCES
Abbasi, A., Chen, H., & Salem, A. (2008). Sentiment analysis in multiple languages: Feature selection for opinion classification in web forums. ACM Transactions on Information Systems, 26(3), 1–34. https://doi.org/10.1145/1361684.1361685
Akcura, T., Altinkemer, K., & Chen, H. (2018). Noninfluentials and information dissemination in the microblogging community. Information Technology and Management, 19(2), 89–106.
Anders, F., Hlawitschka, M., & Fuchs, M. (2020). Automatic classification of infant vocalization sequences with convolutional neural networks. Speech Communication, 119, 36–45.
Angst, C. M., & Agarwal, R. (2009). Adoption of electronic health records in the presence of privacy concerns: The elaboration likelihood model and individual persuasion. *MIS Quarterly, 33*(2), 339–370. https://doi.org/10.2307/20650295

Baum, M., Ognyanova, K., Lazer, D., Volpe, J. D., & Santillana, M. (2020). The State of the Nation: A 50-State Covid-19 Survey. Report #14: Misinformation and Vaccine Acceptance, USA, September 2020.

Carrera, B., & Jung, J. Y. (2018). SentiFlow: An information diffusion process discovery based on topic and sentiment from online social networks. *Sustainability, 10*(8), 2731. https://doi.org/10.3390/su10082731

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology, 39*(5), 752–766. https://doi.org/10.1037/0022-3514.39.5.752

Chernozhukov, V., Kasahara, H., & Schrimpf, P. (2020). Causal impact of masks, policies, behavior on early covid-19 pandemic in the US. *Journal of Econometrics, 220*(1), 23–62. https://doi.org/10.1016/j.jeconom.2020.09.003

Cheung, C. M., & Thadani, D. R. (2012). The impact of electronic word-of-mouth communication: A literature analysis and integrative model. *Decision Support Systems, 54*(1), 461–470. https://doi.org/10.1016/j.dss.2012.06.008

Chris, N., Oliver, B.-B., & Hilde, V. D. B. (2002). *The Media Book*. Hodder Arnold.

Cocosila, M., & Igonor, A. (2015). How important is the “social” in social networking? A perceived value empirical investigation. *Information Technology & People, 28*(2), 366–382. https://doi.org/10.1108/ITP-03-2014-0055

Fan, R., Varol, O., Varamesh, A., Barron, A., van de Leemput, I. A., Scheffer, M., & Bollen, J. (2019). The minutenode dynamics of online emotions reveal the effects of affect labeling. *Nature Human Behaviour, 3*(1), 92–100. https://doi.org/10.1038/s41562-018-0490-5

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

Frank, M. R., Mitchell, L., Dodds, P. S., & Danforth, C. M. (2013). Happiness and the patterns of life: A study of geolocated tweets. *Scientific Reports, 3*(1), 1–9. https://doi.org/10.1038/srep02625

Gao, Q., Tian, Y., & Tu, M. (2015). Exploring factors influencing chinese user’s perceived credibility of health and safety information on weibo. *Computers in Human Behavior, 45*, 21–31. https://doi.org/10.1016/j.chb.2014.11.071

Gigerenzer, G. (1996). On narrow norms and vague heuristics: A reply to Kahneman and Tversky. *Psychological Review, 103*, 592–596. https://doi.org/10.1037/0033-295X.103.3.592

Guo, X., Yu, Z., & Zhou, X. (2014). Predicting the content dissemination trends by repost behavior modeling in mobile social networks. *Journal of Network and Computer Applications, 42*, 197–207. https://doi.org/10.1016/j.jnca.2014.01.015
Neuman, W. L. (1997). Social research methods: Qualitative and quantitative approaches. Allyn & Bacon.

O’Keefe, D. (2015). Persuasion and social influence. Northwestern University: SAGE Publications Inc.

Ormond, D., & Warkentin, M. (2015). Is this a joke? The impact of message manipulations on risk perceptions. *Journal of Computer Information Systems, 55*(2), 9–19. https://doi.org/10.1080/08874417.2015.11645752

Pei, J., Yu, G., Tian, X., & Donnelley, M. R. (2017). A New method for early detection of mass concern about public health issues. *Journal of Risk Research, 20*(4), 516–532. https://doi.org/10.1080/13669877.2015.1100655

Raileanu, L. E., & Stoffel, K. (2004). Theoretical comparison between the gini index and information gain criteria. [In English] *Annals of Mathematics and Artificial Intelligence, 41*(1), 77–93. https://doi.org/10.1023/B:AMAI.0000018580.96245.c6

Simon, H. A. (1954). Bandwagon and underdog effects and the possibility of election predictions. *Public Opinion Quarterly, 18*(3), 245–253.

Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics, 69*, 99–118. https://doi.org/10.2307/1884852

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. In P. Diamond, & M. Rothschild (Eds.), *Uncertainty in Economics* (pp. 17–34). Academic Press.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211*(4481), 453–458. https://doi.org/10.1126/science.7455683

Ucar, T., Culpan, M., Caskurlu, T., Karaman, M. İ., & Silay, M. S. (2018). The activity and discussion points of # Circumcision through Twitter: A microblogging platform. *International Journal of Impotence Research, 30*(5), 249–252. https://doi.org/10.1038/s41443-018-0058-y

Wang, J., Zhou, Y., Zhang, W., Evans, R., & Zhu, C. (2020). Concerns expressed by chinese social media users during the COVID-19 pandemic: Content analysis of sina weibo microblogging data. *Journal of Medical Internet Research, 22*(11), e22152.

Wiederhold, B. K. (2020). Using social media to our advantage: Alleviating anxiety during a pandemic. *Cyberpsychology, Behavior, and Social Networking, 23*(4), 197–198. https://doi.org/10.1089/cyber.2020.29180.bkw

Wu, B., & Shen, H. (2015). Analyzing and predicting news popularity on Twitter. *International Journal of Information Management, 35*(6), 702–711.

Zhao, Y. Y., Qin, B., Liu, T., & Tang, D. Y. (2016). Social sentiment sensor: A visualization system for topic detection and topic sentiment analysis on microblog. *Multimedia Tools and Applications, 75*(15), 8843–8860. https://doi.org/10.1007/s11042-014-2184-y

Zheng, S., Wang, J., Cong, S., Zhang, X., & Kahn, M. E. (2019). Air pollution lowers Chinese urbanites’ expressed happiness on social media. *Nature Human Behaviour, 3*(3), 237–243. https://doi.org/10.1038/s41562-018-0521-2

How to cite this article: Pei, J., Lu, Z., & Yang, X. (2021). What drives people to repost social media messages during the COVID-19 pandemic? Evidence from the Weibo news microblog. *Growth and Change, 00*, 1–18. https://doi.org/10.1111/grow.12573