Postprint: Effectiveness Evaluation and Empirical Study of Online Rumor Debunking During Public Health Emergencies

Authors: Wang Xiwei, Hongfei Zhu, Li Yueqi, He Jinchang

Date: 2023-04-01T16:02:56+00:00

Abstract

[Purpose/Significance] During public health emergencies, the proliferation of online rumors generates adverse impacts on society and people's lives. Constructing a multi-dimensional evaluation index system for rumor-refutation effectiveness can provide guidance for online rumor governance during public health emergencies in China. [Method/Process] This study constructs a rumor-refutation effectiveness evaluation index system from four dimensions: debunking source characteristics, debunking message characteristics, debunking audience characteristics, and debunking background characteristics. Indicator weights are analyzed through a combined method of Analytic Hierarchy Process (AHP) and fuzzy comprehensive evaluation. Empirical analysis is conducted using representative event topics from the Weibo debunking platform during the early stage of COVID-19 to verify the feasibility of the evaluation index system. [Results/Conclusions] Based on empirical research findings from "rumor-refutation Weibo posts," this study derives three strategies for enhancing debunking effectiveness: selecting authoritative and objective debunking sources and implementing collaborative debunking; emphasizing the media framing of debunking information; and conducting psychological and awareness guidance for the debunking audience.

Full Text

Preamble

Evaluation and Empirical Study on the Effectiveness of Online Rumor Refutation During Public Health Emergencies

Wang Xiwei[1],[2],[3], Zhu Hongfei[2], Li Yueqi[2], He Jinchang[2]

[1]School of Management, Jilin University, Changchun 130022
[2]Research Center for Big Data Management, Jilin University, Changchun 130022
[3]Research Center of Cyberspace Governance, Jilin University, Changchun 130022

Abstract: [**Purpose/Significance**] During public health emergencies, the proliferation of online rumors adversely affects society and people's lives. Constructing a multi-angle evaluation index system for refutation effectiveness can provide guidance for online rumor governance in China's public health emergencies. [**Method/Process**] This study constructs an evaluation index system from four dimensions: refutation subject characteristics, refutation text features, audience characteristics, and contextual background features. The Analytic Hierarchy Process (AHP) combined with fuzzy comprehensive evaluation is used to analyze index weights. Empirical analysis is conducted using representative topics from Weibo's rumor-refutation platform during the early COVID-19 period to verify the feasibility of the evaluation system. [**Result/Conclusion**] Based on empirical findings from "refutation Weibo" cases, three strategies for improving refutation effectiveness are proposed: selecting authoritative and objective refutation subjects with collaborative refutation, emphasizing media framework design for refutation information, and guiding audience psychological awareness.

Keywords: public health emergency; online rumor; refutation effectiveness; evaluation index
Classification Number: G250
DOI: 10.13266/j.issn.0252-3116.2021.19.004

The COVID-19 outbreak in 2020 became another major public health emergency in China following the 2003 SARS crisis, significantly impacting people's work and daily lives [1]. During the early outbreak, numerous online rumors, combined with public knowledge gaps about COVID-19, intensified psychological panic. Chinese government agencies and news media continuously refuted rumors to clarify facts and prevent adverse effects. However, online rumors during public health emergencies exhibit a "long-tail" phenomenon after active refutation efforts, with rumor risks persisting even after intervention [2], sometimes leading to refutation failure. Additionally, incorrect refutation methods may trigger a "backfire effect," where the public becomes more convinced of the rumors [3]. Therefore, objectively evaluating refutation effectiveness to enable timely and correct rumor refutation represents an urgent challenge.

This study addresses three research questions: How to construct an evaluation index system for online rumor refutation effectiveness? How to conduct empirical research on refutation effectiveness based on the constructed index system? How can evaluation indicators guide online rumor governance during public health emergencies? By constructing a multi-angle evaluation index system and conducting empirical research on Weibo rumor refutation during COVID-19, this paper validates the effectiveness of the indicators and proposes corresponding guidance for rumor governance in public health emergencies.

This research is supported by the National Social Science Fund Major Project "Research on Big Data-Driven Thematic Graph Construction and Regulation Strategies for Social Network Public Opinion" (Project No.: 18ZDA310).

Author Contributions: Wang Xiwei: proposed research questions and framework, revised manuscript; Zhu Hongfei: wrote manuscript, collected and analyzed data; Li Yueqi: assisted with data processing; He Jinchang: assisted with data collection and analysis.

---

2 Literature Review

2.1 Online Rumors

Online rumors are rumors disseminated through internet platforms, characterized by high attention, wide reach, rapid spread, and easy distortion [11], making them prone to causing negative social impacts. As the internet evolves, online rumor types continue to proliferate, shifting research focus from propagation modeling to analyzing how different rumor content affects the public. Among these, "fear-based rumors" classified by outcome are more likely to attract public attention and spread [12]. Research also shows that among "conspiracy, expectation, and gossip" types categorized by user decision-making behavior, conspiracy rumors have greater impact and cause more harm [13-14].

Online rumors include not only "deliberately fabricated" information but also inaccurate information spread by ordinary users who "believe it to be true" [15], necessitating analysis of refutation timeliness, reach, and effectiveness.

2.2 Online Rumor Refutation

Rumor refutation typically employs three methods: publishing authoritative scientific information to debunk false rumors, countering online rumors through information confrontation, and issuing clarification statements to reduce negative impacts [16]. With deeper research on specific rumor types, refutation studies have become increasingly multi-disciplinary and multi-perspective. Existing research examines refutation information carriers, strategy selection, and subject comparison [6,17-18], offering classifications and recommendations. Interdisciplinary approaches from psychology, management, and journalism analyze social network structures and psychological motivations of communicators and receivers [6], refutation mechanisms [20], and information content [6]. Current research on rumor classification and refutation methods is relatively comprehensive, showing multi-angle, multi-disciplinary trends.

2.3 Rumor Refutation in Public Health Emergencies

Public health emergencies refer to sudden incidents causing or potentially causing serious public health harm, including major infectious disease outbreaks,

unexplained group illnesses, major food and occupational poisoning, and other events severely affecting public health [15]. Due to their significance and suddenness, public health emergencies create structural pressures, information asymmetry, and fear, resulting in rumors with wider propagation, more concentrated sources, and greater potential for malicious exploitation [15]. Current research focuses on analyzing spatiotemporal distribution, channels, and content characteristics of health-related rumors, proposing governance strategies. However, universal strategies are scarce, with more studies based on models for detecting public health rumors [21] or case analyses of different response strategies. In recent years, developing refutation measures based on propagation characteristics has become an academic and institutional focus [22], though research on post-intervention effectiveness evaluation remains limited.

---

3 Evaluation Indicators and Methods

3.1 Evaluation Index System Construction

To systematically evaluate online rumor refutation effectiveness, this study draws upon theoretical frameworks of online refutation [23] regarding mode selection and how refutation subjects convey information to audiences. Four components are extracted: refutation subject characteristics, refutation text features, audience characteristics, and contextual background features.

Through literature review and the refutation implementation process, a hierarchical model for evaluating refutation effectiveness during public health emergencies is established, with refutation effectiveness as the target layer. The criterion layer comprises four dimensions (subject, text, audience, background), and the index layer contains 25 indicators (see Table 1).

3.2 Method Selection and Rationale

AHP and fuzzy comprehensive evaluation are widely used for qualitative indicators, such as evaluating WeChat platform user satisfaction [34] and mobile APP privacy leakage risk [35].

AHP solves qualitative judgments and complex decisions by decomposing problems into hierarchical factors (target, criterion, and indicator layers). Experts compare indicators using Saaty's 1-9 scale to construct pairwise judgment matrices, establishing hierarchical structures. Fuzzy comprehensive evaluation, based on fuzzy mathematics, synthesizes fuzzy relations to quantify boundary-uncertain factors and comprehensively evaluate multiple factors. The combination effectively addresses the strong subjectivity and lack of quantitative indicators in refutation effectiveness evaluation, such as social network structures and emotional orientation [36].

3.3 AHP Weighting Indicators

The consistency ratio is calculated as:

$$CR = CI/RI \quad \text{(Formula 1)}$$

where CR is the random consistency ratio, $CI = (\lambda_{max} - n)/(n-1)$ is the general consistency index, and RI is the average random consistency index. A $CR < 0.1$ indicates acceptable consistency [37].

Five experts in public health emergency public opinion and two rumor audiences (numbered 1-5) were invited to construct judgment matrices. Using the geometric mean method for normalization, the weight vector for matrix M was calculated as [0.1175, 0.2622, 0.5650, 0.0553]. All CR values were below 0.1, passing consistency tests. Final synthetic weights are shown in Table 2 .

---

4 Empirical Analysis

4.1 Weibo Refutation Text Selection

During the early COVID-19 outbreak in January 2020, rumors such as "Banlangen and vinegar can prevent novel coronavirus," "Garlic and saline can kill COVID-19," and "Spring Festival travel canceled, early holiday shutdown" caused widespread panic. Corresponding refutations emerged on Weibo, which became the primary platform for information acquisition and discussion. The "Weibo Refutation" official account, verified for credibility and objectivity [26], provided representative samples. Three top-ranked refutation topics by forwarding and comment volume were selected: People's Daily's "Spread! #Nine Latest Epidemic Rumors#, Don't Be Fooled," CCTV News's "Don't Believe or Spread! #Epidemic Refutation Information Summary#," and Communist Youth League Central Committee's "#Refutation! These Rumors Are All False#." Details are in Table 3 .

4.2 Rating Process

The fuzzy comprehensive evaluation procedure involves: establishing factor sets from the index system; determining evaluation comment sets. The comment set is defined as $V = \{\text{very poor, poor, average, good, very good}\}$; constructing membership matrices. Due to qualitative indicators like user relationships and emotions, 15 network public opinion experts rated the four dimensions to obtain fuzzy relation matrices; calculating comprehensive results. Using the weighted average principle, comment grades correspond to scores $\{s_1, s_2, s_3, s_4, s_5\} = \{60, 70, 80, 90, 100\}$ for quantification (see Table 4).

Taking "refutation subject appropriateness" as an example: 0% rated it as poor or very poor, 6.67% as average, 40% as good, and 53.3% as very good, yielding the membership vector $r_{11} = \{0, 0, 0.0667, 0.4, 0.533\}$. Repeating this for

all indicators under "refutation subject characteristics" constructs membership matrix R_1, and similarly for other dimensions.

Using AHP-derived weight vectors W_i and membership matrices R_i with multiplication operators, the evaluation result for refutation subjects C_1 is obtained. Combining four-dimensional results yields fuzzy matrix C_B. The final result is:

$$C = WC_B = (0.1175, 0.2622, 0.5650, 0.0553) \times \begin{bmatrix} 0 & 0.0118 & 0.1248 & 0.2074 & 0.6556 \\ 0 & 0.1719 & 0.2647 & 0.5260 & 0.1923 \\ 0 & 0.0352 & 0.3590 & 0.4000 & 0.2065 \\ 0.2732 & 0 & 0 & 0.2306 & 0.2608 \end{bmatrix} = (0, 0.0252, 0.2985,$$

With comment standard values $S = (60, 70, 80, 90, 100)$, the weighted score $p = C \cdot S^T = 89.13$, indicating "good" effectiveness.

4.3 Results

4.3.1 Primary Indicator Weights

"Audience characteristics" scored highest (0.5650), indicating it most influences refutation effectiveness. Regardless of subject or media, the ultimate impact is on users. Refutation aims to help the public distinguish rumors and avoid irrational behavior. Other dimensions ranked: "text features" (0.2622), "subject characteristics" (0.1175), and "background features" (0.0553). The low weight for background features suggests that mode, channel, and strategy selection have limited overall impact.

4.3.2 Secondary Indicator Weights

Top secondary indicators were "user centrality" (0.1965), "user attention" (0.1418), and "user acceptance" (0.0913), showing that users closely related to events pay more attention and are more easily persuaded. "Refutation timeliness" (0.0774) and "subject authority" (0.0709) ranked fourth and fifth, demonstrating that timely release prevents rumor spread, and official media refutation gains widespread recognition. "Refutation strategy selection" ranked lowest (0.0024), indicating audiences are familiar with combined debunking and clarification approaches.

4.3.3 Overall Effectiveness Evaluation

The comprehensive evaluation score of 89.13 indicates users were basically satisfied but see room for improvement. "Subject characteristics" (95.27) and "background features" (90.92) scored high, showing government-media-opinion leader hierarchical models have strong authority and recognition. However, "text features" (88.95) and "audience characteristics" (87.77) scored lower despite higher weights, indicating current measures are inadequate. While user acceptance reached 94, attention, centrality, and interaction frequency were around 86, suggesting users passively accept information without becoming active disseminators.

5 Discussion and Analysis

5.1 Subject Selection and Collaborative Refutation

With improved mechanisms, multiple subjects should form systematic collaboration. Based on authoritative government media, multi-role subjects should establish hierarchical relationships to expand social networks and guide real-time information release. Subjects can create shared topics, with government leading media and opinion leaders to establish hierarchical interactions, ultimately reaching ordinary users. This ensures credibility while improving dissemination efficiency.

5.2 Media Framework Design for Refutation Information

Current refutation information often appears as lengthy, static text, inhibiting dissemination willingness. Results show media framework significantly affects effectiveness. Users prefer "fragmented" reading with minimal time investment. Effective refutation combines text, audio, and video to present facts conveniently [26]. Multi-dimensional presentation completes media framework setting. Additionally, text emotional orientation scored low; during public health emergencies, negative emotions are common. Refutation content should be concise, focus on core issues, objectively state facts, and convey positive emotions [26] to encourage attention and sharing.

5.3 Guiding Audience Psychological Awareness

Audiences often selectively receive information, adopting a "better safe than sorry" attitude toward rumors, hindering refutation effectiveness [30]. Users attribute rumor spread to social interaction motives, trusting rumor spreaders while refusing to disseminate refutations. Though audiences highly trust authoritative sources, they remain passive recipients. As the highest-weight indicator, audience characteristics determine success. Low attention and participation highlight the importance of psychological guidance. Refutation subjects should promote user willingness to spread by emphasizing altruism and self-improvement, using text to guide participation and forwarding. Timely release during outbreaks is crucial for guiding and reassuring audiences.

6 Research Conclusions

This study theoretically constructs a four-dimensional evaluation index system for public health emergency rumor refutation effectiveness, using AHP and fuzzy evaluation to determine weights. Empirically, representative COVID-19 refutation Weibo cases validate the system's scientific feasibility. Based on the "good"

refutation effectiveness, three improvement strategies are proposed: authoritative subject selection with collaborative refutation, media framework design, and audience psychological guidance.

Limitations include: subjective user scoring may cause errors; limited coverage using only Weibo cases. Future research will expand to more platforms and use quantitative methods to improve generalizability.

References

[1] iMedia Consulting. "2020 China COVID-19 Positive/Negative Impact Industries and Typical Enterprise Performance Impact Analysis Report" [EB/OL]. [2020-02-24]. https://www.iimedia.cn/c400/69270.htm.
[2] Zhang Guirong, Xia Ting. Research on the Long-tail Effect Control of Online Rumor Propagation in Public Emergencies [J]. Information Studies: Theory & Application, 2021, 44(3): 69-75.
[3] Xiong Yan. Explanation and Suppression Strategies for the Backfire Effect in Rumor Propagation: Integration and Derivation Based on Empirical Research [J]. Modern Communication (Journal of Communication University of China), 2019, 41(1): 75-81.
[4] Hosni A, Li K. Minimizing the influence of rumors during breaking news events in online social networks [J]. Knowledge-based systems, 2020, 193: 105452.
[5] Nan Z, Hong H, Boni S, Jing LZ, Bo Z. Dynamic 8-state ICSAR rumor propagation model considering official rumor refutation [J]. Physica A: statistical mechanics and its applications, 2014, 415: 333-346.
[6] Tian RY, Zhang XF, Liu YJ. SSIC model: A multi-layer model for intervention of online rumors spreading [J]. Physica A: statistical mechanics and its applications, 2015, 427: 181-191.
[7] Ma Chao. Research on the Category Composition and Community Structure of Health Topic Refutation Communities: From the Perspective of Multi-agent Rumor Collaborative Governance [J]. Journal of Intelligence, 2019, 38(1): 96-105.
[8] Qu Nanwei, Xia Zhijie, Wang Yiming. Research on Social Media Refutation Effectiveness Based on User Information Behavior [J]. Information Science, 2021, 39(1): 111-119.
[9] Chen Juan, Liu Yanping, Deng Shengli. Research on Factors Influencing User Comments and Emotional Tendencies of Government Refutation Information [J]. Information Science, 2017, 35(12): 61-65.
[10] Jin Yalan, Xu Qizhi. Research on Public Participation-based Rumor Refutation Mechanisms in Public Health Emergencies: Taking "Ding Xiang Doctor" and "Science China" as Examples [J]. Science Popularization, 2020, 15(2): 52-59, 106.
[11] Tong Ranran. Literature Review on Public Health Event Online Rumors [J]. Media Forum, 2019, 2(20): 159-160.

[12] Cui Jindong, Chen Siyuan, Li Chenyu. Empirical Research on Multi-type Online Rumor Quelling Methods Based on Big Data: Taking "Rumors During COVID-19" as an Example [J/OL]. Information Studies: Theory & Application, 2021(4): 1-15.
[13] Zhang Xing, Wu You, Xia Huosong. Research on Online Health Rumor Propagation Willingness: The Role of Rumor Source, Type, and Propagation Object [J]. Nankai Business Review, 2020, 23(1): 200-212.
[14] Li CC, Liu FM, Li P. Ising model of user behavior decision in network rumor propagation [J]. Discrete dynamics in nature and society, 2018, 18(9): 122-135.
[15] Cai Xu. Governance of Online Rumors in Public Health Events in the All-media Era: Taking COVID-19 Online Rumor Governance as an Example [J]. Journal of Xiamen Special Zone Party School, 2021(1): 33-36.
[16] Tang Xuemei, Lai Shengqiang. Research on Government Online Rumor Refutation Strategies in Emergencies: Taking the Taifu Middle School Incident as an Example [J]. Journal of Intelligence, 2018, 37(9): 95-100.
[17] Wang Guohua, Wang Ge, Yang Tengfei, et al. Research on the Operation and Effect of Online Rumor Refutation Platforms [J]. Journal of Intelligence, 2014, 33(9): 100-105, 134.
[18] Du Ze, Zhang Xiaojie. Research on Online Public Opinion Governance in Public Health Emergencies from the Perspective of Evidence-based Governance [J]. Information Studies: Theory & Application, 2020, 43(5): 17-23.
[19] Wang Chenyu. Problems and Suggestions for Online Rumor Refutation Information Propagation from a Psychological Perspective [J]. Journal of News Research, 2018, 9(17): 50-52, 123.
[20] Xu Jiao. Generation Mechanism and Social Structure Governance of Online Rumors [J]. Public Communication of Science & Technology, 2019, 11(9): 127-128.
[21] Santhoshkumar S, Babu D. Earlier detection of rumors in online social networks using certainty-factor-based convolutional neural networks [J]. Social network analysis and mining, 2020, 10(4): 993-1022.
[22] Paek HJ, Hove T. Effective strategies for responding to rumors: the case of radiation-contaminated food in South Korea [J]. Public relations review, 2019, 45(3): 25-37.
[23] Tong Wensheng, Yi Baihui. Online Rumor Refutation: Domestic Research Progress and Theoretical Analysis Framework [J]. Journal of Intelligence, 2020, 39(6): 128-134, 202.
[24] Tang Mengfei, Wang Jiancheng. Research on Government Weibo Rumor Refutation Effectiveness in Emergencies: Based on the "Shanghai Bund Stampede" Case Analysis [J]. Journal of Intelligence, 2015, 34(8): 98-103, 36.
[25] Chen Juan, Liu Yanping, Deng Shengli. Research on Factors Influencing User Comments and Emotional Tendencies of Government Refutation Information [J]. Information Science, 2017, 35(12): 61-65.
[26] Li Zongmin, Zhang Qi, Du Xinyu. Research on Refutation Effectiveness Based on Interactive and Popular Comment Emotional Tendencies of Refutation Weibo: Taking COVID-19 Related Refutation Weibo as an Example [J].

Journal of Intelligence, 2020, 39(11): 90-95, 110.
[27] Li ZM, Zhang Q, Du XY, et al. Social media rumor refutation effectiveness: evaluation modeling and enhancement [J]. Information processing & management, 2021, 58(1): 102420.
[28] Qiang Yuexin, Sun Zhipeng. Research on Government Weibo Rumor Refutation Effectiveness from the Perspective of Political Communication [J]. Journalism University, 2020, (10): 1-15, 118.
[29] Wang C, Huberman BA. Long trend dynamics in social media [J]. EPJ data science, 2012, 1(1): 2.
[30] Wang Chao. Why Does Rumor Refutation Fail? An Interpretive Framework from the Perspective of Information Propagation Effects [J]. Journal of Intelligence, 2019, 38(5): 123-129.
[31] Freeman LC. Centrality in social networks conceptual clarification [J]. Social networks, 1978, 1(3): 215-239.
[32] Zong Qianjin, Huang Zifeng. Research on the Diffusion Mechanism of Social Media Rumor Refutation Information: Based on Longitudinal Analysis of Refutation Information Evolution Networks [J]. Modern Intelligence, 2017, 37(9): 44-52, 121.
[33] Wang Guohua, Wang Ge, Yang Tengfei, et al. Research on the Operation and Effect of Online Rumor Refutation Platforms [J]. Journal of Intelligence, 2014, 33(9): 100-105, 134.
[34] Yang Shaomei, Wang Ting, Li Shengli. Comprehensive Evaluation of WeChat English Learning Platform User Satisfaction Based on Fuzzy Analytic Hierarchy Process: Taking Shuǐdī Reading as an Example [J]. Library and Information Service, 2019, 63(21): 97-104.
[35] Tian Bo, Zheng Yusha, Liu Pengyuan, et al. Research on Mobile APP User Privacy Information Leakage Risk Evaluation Indicators and Empirical Study [J]. Library and Information Service, 2018, 62(19): 101-110.
[36] Zou Wenwu. Research on the Development Dynamics and Knowledge Graph of Social Network User Influence [J]. Information Science, 2020, 38(9): 107-115.
[37] Du Dong, Pang Qinghua, Wu Yan. Modern Comprehensive Evaluation Methods and Case Selection [M]. Beijing: Tsinghua University Press, 2008: 9-33.

Evaluation and Empirical Study on the Effectiveness of Internet Rumor Refutation During Public Health Emergencies

Wang Xiwei[1],[2],[3], Zhu Hongfei[2], Li Yueqi[2], He Jinchang[2]

[1]School of Management, Jilin University, Changchun 130022
[2]Research Center for Big Data Management, Jilin University, Changchun 130022
[3]Research Center of Cyberspace Governance, Jilin University, Changchun 130022

Abstract: [**Purpose/Significance**] During public health emergencies, numerous internet rumors emerge, adversely affecting society and people's lives. Constructing a multi-perspective evaluation index system for refutation effectiveness can provide guidance for internet rumor governance in China during such emergencies. [**Method/Process**] This paper builds an evaluation index system from four dimensions: refutation subject characteristics, refutation text features, audience characteristics, and contextual background features. The Analytic Hierarchy Process (AHP) combined with fuzzy comprehensive evaluation analyzes index weights. Empirical analysis verifies the system's feasibility using representative topics from Weibo's rumor-refutation platform during the early COVID-19 period. [**Result/Conclusion**] Based on empirical research of "refutation Weibo," three strategies are proposed: selecting authoritative, objective refutation subjects with collaborative refutation; emphasizing media framework design for refutation information; and guiding audience psychological awareness.

Keywords: public health emergency; internet rumor; refutation effectiveness; evaluation index

Note: Figure translations are in progress. See original paper for figures.